



## OPEN ACCESS

## EDITED BY

Linchun Shi,  
Institute of Medicinal Plant  
Development, Chinese Academy of  
Medical Sciences and Peking Union  
Medical College, China

## REVIEWED BY

Wenjun Zhu,  
Wuhan Polytechnic University, China  
Wei Sun,  
China Academy of Chinese Medical  
Sciences, China

## \*CORRESPONDENCE

Yinglin Liu  
✉ lyldaliedu@163.com  
Ping Zhou  
✉ zhouping725@126.com

## SPECIALTY SECTION

This article was submitted to  
Plant Bioinformatics,  
a section of the journal  
Frontiers in Plant Science

RECEIVED 16 November 2022

ACCEPTED 13 December 2022

PUBLISHED 06 January 2023

## CITATION

Fang H, Dai G, Liao B, Zhou P and  
Liu Y (2023) Application of chloroplast  
genome in the identification of  
*Phyllanthus urinaria* and its  
common adulterants.  
*Front. Plant Sci.* 13:1099856.  
doi: 10.3389/fpls.2022.1099856

## COPYRIGHT

© 2023 Fang, Dai, Liao, Zhou and Liu.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Application of chloroplast genome in the identification of *Phyllanthus urinaria* and its common adulterants

Hui Fang, Guona Dai, Binbin Liao, Ping Zhou\* and Yinglin Liu\*

College of Pharmaceutical Science, Dali University, Dali, China

**Background:** *Phyllanthus urinaria* L. is extensively used as ethnopharmacological material in China. In the local marketplace, this medicine can be accidentally contaminated, deliberately substituted, or mixed with other related species. The contaminants in herbal products are a threat to consumer safety. Due to the scarcity of genetic information on *Phyllanthus* plants, more molecular markers are needed to avoid misidentification.

**Methods:** In this study, the complete chloroplast genome of nine species of the genus *Phyllanthus* was *de novo* assembled and characterized.

**Results:** This study revealed that all of these species exhibited a conserved quadripartite structure, which includes a large single copy (LSC) region and small single copy (SSC) region, and two copies of inverted repeat regions (IRa and IRb), which separate the LSC and SSC regions. And the genome structure, codon usage, and repeat sequences were highly conserved and showed similarities among the nine species. Three highly variable regions (*trnS-GCU-trnG-UCC*, *trnT-UGU-trnL-UAA*, and *petA-psbJ*) might be helpful as potential molecular markers for identifying *P. urinaria* and its contaminants. In addition, the molecular clock analysis results showed that the divergence time of the genus *Phyllanthus* might occur at ~ 48.72 Ma.

**Conclusion:** This study provides valuable information for further species identification, evolution, and phylogenetic research of *Phyllanthus*.

## KEYWORDS

*Phyllanthus urinaria*, chloroplast genome, species identification, molecular marker, phylogenetic

## 1 Introduction

*Phyllanthus urinaria* L. belongs to the family Euphorbiaceae, is listed in the dictionary of Chinese ethnic medicine, and the Chinese name is “Yexiazhu” (Guo et al., 2017). The herbal *P. urinaria* has crucial medicinal value in anti-diarrheal, anti-inflammatory, jaundice, diabetes, malaria, hepatitis B, and liver diseases (Chudapongse et al., 2010; Geethangili and Ding, 2018). The previous survey has revealed that *P. urinaria* is generally contaminated with other common adulterants, such as *P. acidus* (L.) Skeel, *P. amarus* Schumacher & Thonning, *P. reticulatus* Poir., *P. niruri* L., *P. emblica* L., *P. pulcher* Well. ex Muell. Arg., and *P. debilis* Klein ex Willd. (Manissorn et al., 2010; Srirama et al., 2010; Kiran et al., 2021). These adulterants are usually of poor quality, and some might even be toxic (Adedapo et al., 2005; Geng et al., 2021). As the morphology of these species are interchangeable, similar, and indistinguishable, the identification of these species remains controversial, which may impair their clinical safety and efficacy (Sarin et al., 2013; Kiran et al., 2021). Therefore, it is essential to develop a method for accurately identifying *P. urinaria* and its common contaminants.

With the rapid development of molecular technology, molecular identification has made significant progress in Chinese medicine, especially molecular markers, a technique that involves sequencing specific sections of the genome to identify differences between individuals of different species or populations (Xiong et al., 2018). Recent studies have revealed high levels of genetic diversity and a lack of phylogenetic resolution within species of *Phyllanthus* (Pruesapan et al., 2012; Bouman et al., 2021). Universal DNA barcodes, such as ITS, *psbA-trnH*, *trnL*, *psbK-psbI*, *rpoC1*, and *trnL-trnF*, have been used to identify *P. urinaria* and its adulterants (Manissorn et al., 2010; Srirama et al., 2010; Inglis et al., 2018; Kiran et al., 2021). However, some common adulterants were not included in these investigations, and there are inherent limitations to single-locus DNA barcodes (Heinze, 2007; Li et al., 2015). Therefore, more scientific and accurate identification methods must be developed. The chloroplast (cp) is an essential organelle that plays a crucial role in plant photosynthesis and several other critical biochemical processes (Neuhaus and Emes, 2000). Compared with the traditional DNA fragments, the cp genome was relatively conserved and slightly varied, which has been applied to many research fields, such as species identification and the development of molecular markers (Abdullah et al., 2020; Li et al., 2022; Wang et al., 2022). The method has been widely used for identifying *Paris*, *Polygonatum*, *Vicatia*, and their adulterants (Guan et al., 2022; Jiang et al., 2022; Wang et al., 2022). Recently, although the complete cp genomes of Phyllanthaceae species have been reported and the high-resolution phylogenetic tree was reconstructed (Rehman et al., 2021), the purpose of this study was to clarify the genome evolution in Phyllanthaceae and identify the polymorphic loci

for phylogenetic inference. To our knowledge, no reports use cp genomes to compare *P. urinaria* with its common adulterants.

Our study aims to: (i) contribute new fully-sequenced cp genomes of *Phyllanthus* and improve the understanding of the overall structure of these genomes, (ii) perform comparative analyses and elucidate the phylogenetic evolution of the *Phyllanthus*, and (iii) screen potential molecular markers to distinguish *P. urinaria* from its contaminants. In the current work, the complete cp genomes of nine *Phyllanthus* species were sequenced, *de novo* assembled, and annotated. These genomes were then used in a comparative analysis of genome structure and evolution relationships. This research expands the genomic resources available for *Phyllanthus* and provides valuable information support for the phylogenetic analysis and identification of the *Phyllanthus*, as well as for the safe applications of *P. urinaria*.

## 2 Material and methods

### 2.1 Plant and DNA sources

The fresh and healthy leaves for nine species of *P. acidus*, *P. amarus*, *P. reticulatus*, *P. urinaria*, *P. niruri*, *P. niruri* subsp. *lathyroides*, *P. emblica*, *P. pulcher*, and *P. franchetianus* were collected from Dali and Xishuangbanna, Yunnan Province, China. The detailed information per sample is available in Supplementary Table 1. The samples were identified following the taxonomic key and external morphology diagnosis proposed by related literature (Webster and Carpenter, 2008). The voucher specimens were preserved at the herbarium of Dali University. The fresh leaf of nine species was frozen in liquid nitrogen and stored in a 4°C refrigerator for DNA extraction. Total DNA was extracted using a modified cetyl trimethyl ammonium bromide (CTAB) procedure (Allen et al., 2006). DNA quality and quantity were assessed using a NanoDrop spectrophotometer (ND-2000; Thermo Fisher Scientific, USA) and agarose gel electrophoresis.

### 2.2 DNA sequencing, assembly and annotation

Purified high-quality genomic DNA was broken into short fragments of approximately 350 bp, and paired-end (PE) libraries were constructed by adding A-tails, PCR amplification, and other steps, followed by sequencing in 150 bp paired-end mode on an Illumina NovaSeq 6000 platform. The high-quality reads were assembled using GetOrganelle v1.7.5 (Jin et al., 2020) and then annotated by cpGAVAS2 (<http://47.96.249.172:16019/analyzer/annotate>) and (GeSeq, RRID : SCR\_017336) (<https://chlorobox.mpimp-golm.mpg.de/geseq.html>) (Tillich et al., 2017; Shi et al., 2019). The annotations of tRNA genes were confirmed by using

(tRNAscan-SE v.2.03, RRID : SCR\_010835) (<http://lowelab.ucsc.edu/tRNAscan-SE/>) (Schattner et al., 2005). Annotated cp genomes sequences were submitted to GenBank and are available under accession numbers OP009343-OP009351 (Table 1). Fully annotated cp genome circular diagrams were drawn by OrganellarGenomeDRAW (OGDRAW, RRID : SCR\_017337) (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) (Greiner et al., 2019).

## 2.3 Genome structure and comparisons analysis

Forward (F), palindromic (P), reverse (R), and complementary (C) were identified using the REPuter (<https://bibiserv.cebitec.uni-bielefeld.de/reputer/>) tool (Kurtz et al., 2001). The criteria for identifying repeats include a minimum repetition size of 30 bp and a 90% similarity between repeat pairs, calculated by assigning a value of 3 to the altered sequence. In addition, (MISA, RRID : SCR\_010765) (<http://pgrc.ipk-gatersleben.de/misa/>) software was used to identify simple sequence repeats (SSRs) (Beier et al., 2017). We followed conventional standards for identifying chloroplast and mitochondrial SSRs, including a minimum stretch of 10 for

mono-, 5 for di-, 4 for tri-, and 3 for tetra-, penta- and hexanucleotide repeats and a minimum distance of 100 bp between compound SSRs. Relative synonymous codon usage (RSCU) was analyzed by CodonW v.1.4.2 (Sharp et al., 1986). Also, Tltools v1.098761 used a heatmap to show the values of RSCU (Chen et al., 2020).

For comparative analysis of genes, tRNA, repeat content, genome size, and GC content were assessed by (Geneious v.2022.2.1, RRID : SCR\_010519) (<http://www.geneious.com/>) software (Kearse et al., 2012). The software mVISTA (<https://genome.lbl.gov/vista/index.shtml>) (Frazer et al., 2004) in Shuffle-LAGAN mode (Brudno et al., 2003) was used to compare the nine *de novo* cp genome sequences, *P. amarus* (GenBank OP009344) was used as the reference genome. IRscope (<https://irscope.shinyapps.io/irapp/>) was used to analyze inverted repeat region contraction and expansion at the junctions of cp genomes (Amiryousefi et al., 2018). The cp genomes were aligned in (MAFFT, RRID : SCR\_011811) (<https://mafft.cbrc.jp/alignment/server/>). Additionally, the nucleotide variability (Pi) across the cp genome sequences was performed in (DnaSP v.6.12.03, RRID : SCR\_003067) (<http://www.ub.edu/dnasp/>) (Rozas et al., 2017), with a window length of 600 bp and step size of 200 bp. A value of Pi higher than 0.05 was recommended as mutational hotspots (Ren et al., 2022).

TABLE 1 Cp genomes features of nine species of *Phyllanthus*.

Genome features	<i>P. acidus</i>	<i>P. amarus</i>	<i>P. reticulatus</i>	<i>P. urinaria</i>	<i>P. niruri</i>	<i>P. niruri</i> subsp. <i>lathyroides</i>	<i>P. emblica</i>	<i>P. pulcher</i>	<i>P. franchetianus</i>
Genome size (bp)	156,331	155,790	156,610	153,850	155,900	143,563	155,841	155,589	155,598
LSC size (bp)	85,807	85,185	85,868	83,714	85,307	91,305	85,721	85,533	85,533
SSC size (bp)	19,262	19,015	19,182	18,780	19,015	18,986	18,950	18,790	18,799
IRa/IRb size (bp)	25,631	25,795	25,780	25,678	25,789	16,771	25,585	25,633	25,633
Total GC content (%)	36.9	36.6	36.6	36.9	36.6	36.8	36.8	36.8	36.8
GC content in LSC (%)	34.6	34.3	34.3	34.5	34.2	34.9	34.5	34.4	34.4
GC content in S.S.C. (%)	30.6	30.0	30.2	30.6	30.0	30.1	30.4	30.9	30.9
GC content in IRa/IRb (%)	43.1	42.9	42.9	43.0	42.9	45.6	43.1	42.9	42.9
Number of genes	126	125	125	122	125	118	126	123	123
Protein-coding genes	82	81	82	79	81	75	82	79	79
tRNA genes	36	36	35	35	36	35	36	36	36
rRNA genes	8	8	8	8	8	8	8	8	8
Accession numbers in GenBank	OP009343	OP009344	OP009345	OP009346	OP009347	OP009348	OP009349	OP009350	OP009351

## 2.4 Phylogenetic analysis and divergence times analysis

For the phylogenetic analysis, 38 Euphorbiaceae taxa initially consisted of 29 species downloaded from NCBI (Table S2) and the 9 species presented here (Table 1). At the same time, two species, *Daphniphyllum oldhamii* (GenBank NC037883) and *D. macropodum* (GenBank MN496060) were selected as outgroups (Chase et al., 2016). A total of 40 cp genomes were aligned using MAFFT with the default parameters and trimmed using (TrimAl v.1.4, RRID : SCR\_017334) (<http://trimal.cgenomics.org/>) with an automated option (Katoh and Standley, 2013). The best-fit model of nucleotide substitution was selected using ModelFinder (Kalyaanamoorthy et al., 2017) with the Bayesian information criterion as implemented in (IQ-tree v.1.6.12, RRID : SCR\_017254) (<http://www.iqtree.org/>) (Nguyen et al., 2015). The alignment was also evaluated using bootstrap analysis on 1,000 in a maximum likelihood (ML) by IQ-tree v.1.6.12, with the following parameters: `iqtree -s input -m TVM+F+R3 -bb 1000 -alrt 1000 -nt AUTO -o NC_037883, MN_496060`. Besides, the neighbor-joining (NJ) tree was constructed using (MEGA X v.10.2.6, RRID : SCR\_000667) (<http://megasoftware.net/>), and the bootstrap testing was performed with 1,000 repetitions (Kumar et al., 2018).

For analysis of divergence times, the molecular clock tree was constructed based on an ML tree using MEGA X v.10.2.6 (Kumar et al., 2018; Mello, 2018). The relevant divergence times were executed in the (TimeTree, RRID : SCR\_021162) (<http://www.timetree.org/>) Resource (Kumar et al., 2017). Four calibration points were used to restrict each node: (F1) 110.9–121.0 Ma for the root node, (F2) 48.6–55.8 Ma for Phyllanthoideae stem age, (F3) 3.5–74.3 Ma for Acalyphoideae crown age, and (F4) 21.4–89 Ma for Euphorbioideae + Crotonoideae.

## 3 Results

### 3.1 Genome structure

The raw data of nine species were filtered to remove adapters and low-quality reads. Approximately 2.24–4.09 Gb data were obtained for each species. The cp genomes of these nine species are small circular DNA molecules with sizes in the range of 143,563 bp (*P. niruri* subsp. *lathyroides*) to 156,610 bp (*P. reticulatus*) (Figure 1), with the typical quadripartite structure of land plant cp genomes consisting of two inverted repeats (IRA and IRb) separated by large single copy (LSC) and small single copy (SSC) regions, respectively. The size of LSC ranged from 83,714 bp (*P. urinaria*) to 91,305 bp (*P. niruri* subsp. *lathyroides*), SSC ranged from 18,780 bp (*P. urinaria*) to 19,262 bp (*P. acidus*), and the size of each IR region ranged from 16,771 bp (*P. niruri* subsp. *lathyroides*) to 25,795 bp (*P. amarus*). Moreover, the GC content in the IR region (42.9%–

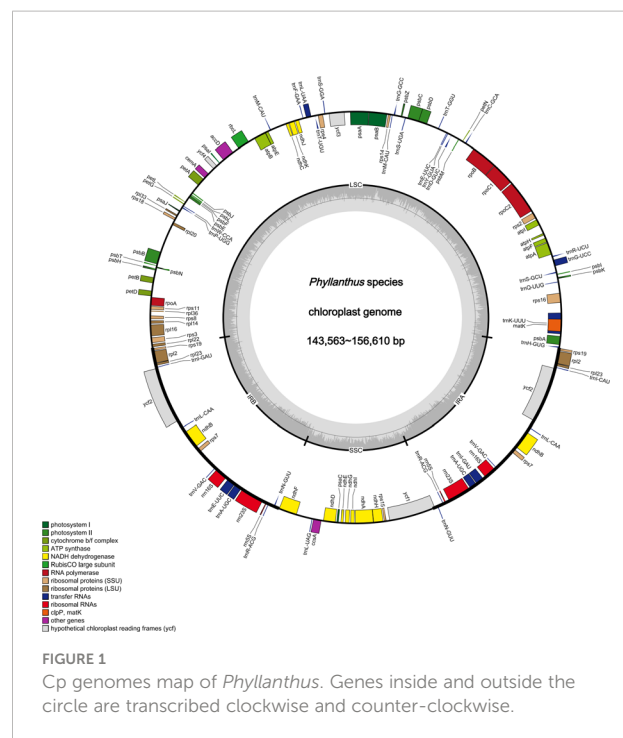


FIGURE 1  
Cp genomes map of *Phyllanthus*. Genes inside and outside the circle are transcribed clockwise and counter-clockwise.

45.6%) was higher than LSC (34.2%–34.9%) and SSC (30.0%–30.9%) (Table 1).

In addition, a total of 118–126 genes were identified, which comprised 75–82 protein-coding genes, 35–36 tRNAs, and 8 rRNAs (Table 1), whereas the number of genes varies in species due to IRs contraction and expansion. These genes were divided into three parts, of which 45 genes belong to photosynthesis-related genes (photosystem I, photosystem II, NADPH dehydrogenase, cytochrome b/f complex ATP synthase, and rubisco), 27 genes belong to self-replication (the large subunit of the ribosome, small subunit of the ribosome, and DNA dependent RNA polymerase), and the remaining genes belong to other genes (acetyl-CoA-carboxylase, c-type cytochrome synthesis genes, envelop membrane proteins, proteases, and maturase) (Table S3). Moreover, 17 genes each contained one intron, among them *rpl2* (×2), *ndhB* (×2), *trnI-GAU* (×2), and *trnA-UGC* (×2), which were located in the IR, and the genes (*trnK-UUU*, *rps16*, *trnG-UCC*, *rpoC1*, *ycf3*, *trnL-UAA*, *trnV-UAC*, and *clpP*) were located in the LSC, while the *ndhA* was only present in the SSC region. In addition, the *ycf3* and *clpP* each contain two introns (Table S3).

### 3.2 Repeat analysis

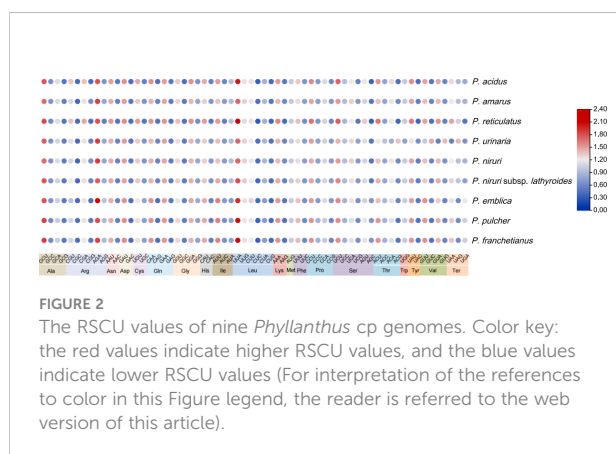
Repetitive sequences in cp genomes play a critical role in genome evolution and rearrangements. Analysis of oligonucleotide repeat revealed that the number of repeat types varied among the nine cp genomes and presented random permutations, and most repeat sequences were within 30–39 bp

(Figure S1). Meanwhile, the frequency of F and P repeats was greater than that of R and C repeats. The structural analysis of the repetition sequence is shown in Figure S1. The minimum number of repeats was found in *P. niruri* subsp. *lathyroides* (26), whereas the maximum was found in *P. niruri* (49).

SSRs, also known as microsatellites, consists of repeating units of 1–6 bp in length. 56, 98, 62, 80, 97, 54, 81, 69, and 69 SSRs were identified in *P. acidus*, *P. amarus*, *P. reticulatus*, *P. urinaria*, *P. niruri*, *P. niruri* subsp. *lathyroides*, *P. emblica*, *P. pulcher*, and *P. franchetianus*, respectively (Table S4). The number of SSRs showed the highest in *P. amarus* (98) and the lowest in *P. niruri* subsp. *lathyroides* (54). Most SSRs were found in LSC regions instead of SSC and IR regions (Figure S2). More than half of the SSRs (51.85%–66.67%) were mononucleotides with the A/T motif, followed by dinucleotides (16.25%–31.48%) with a predominant motif of AT/TA, trinucleotides (1.85%–6.25%) with a predominant motif of AAT/ATT, tetranucleotide repeats (1.61%–4.35%) with a predominant motif of AAAT/ATTT, pentanucleotides (0–1.61%), and hexanucleotides (0–1.25%) that only exist in the cp genome of *P. urinaria*.

### 3.3 Codon usage bias of cp genomes

The analyses of RSCU provide information about the encoding frequency of codons for an amino acid. There were 64 codons in the coding sequence of nine *Phyllanthus* species genes, among which 61 codons encoded 20 amino acids, and the other three codons (UAA, UAG, and UGA) were stop codons (Table S5). Amino acid frequency analyses revealed that the highest frequencies were leucine and isoleucine, whereas cysteine was a rare amino acid. The codon exhibited a strong bias toward an A or T at the third position. An RSCU value below 1.00 indicates that the codon usage frequency is lower than expected, whereas an RSCU value above 1.00 indicates that the codon usage frequency is higher than expected. In this study, the RSCU values of 30 codons were greater than 1, whereas the RSCU value of 32 codons was less than 1, and 2 codons were equal to 1 (Figure 2).



Moreover, the results showed that the GC content of synonymous third codon positions (GC3s) is closely related to codon bias, and the values of GC3s ranged from 25.0% to 31.1%, suggesting that the genus *Phyllanthus* had a greater preference for the A/U ending codons. And the GC content of these cp genomes was highly conserved. In addition, the values for the effective number of codons ranged from 48.43% to 52.93%. Both the codon adaptation index and optimal frequency were less than 0.5. These findings indicated a slight bias toward codon usage in the nine *Phyllanthus* species.

### 3.4 Inverted repeats

Expansion and contraction at the borders of IR regions are common evolutionary phenomena that may explain variations in the size of cp genomes. As illustrated in Figure 3, the *rps3* genes existed entirely in the LSC regions of all species, and *rpl2* existed entirely in the IR region except for *P. niruri* subsp. *lathyroides*. A truncated copy of the *rpl22* gene was observed at the junction of LSC/IRb in three species (*P. reticulatus*, *P. pulcher*, and *P. franchetianus*), which starts in LSC regions and integrates into the IRb region with a size ranging from 2 to 23 bp, whereas the remaining six species were present entire in the LSC region. Another truncated copy of the *rps19* gene was found at the junction of IRA/SSC in two species (*P. acidus* and *P. emblica*). Notably, *rps19* exists entirely in the LSC region for *P. niruri* subsp. *lathyroides*, and *rps19* is present entirely in the IR region in the remaining six species. Besides, the *ndhF* gene was found in the SSC regions except for three species (*P. acidus*, *P. pulcher*, and *P. franchetianus*), which start in the SSC regions and integrate into the IRb region in those species. Moreover, the *ycf1* gene was observed at the IRA/SSC junction except for *P. urinaria*. In all other species, the *ycf1* gene starts in SSC regions and integrates into the IRA. However, in *P. urinaria*, the *ycf1* gene is completely present in the SSC region. Both *psbA* and *trnN* exist entirely in the LSC and IRA, respectively. Notably, the *trnL* gene exists only in the IR region of *P. niruri* subsp. *lathyroides*. These results show that the cp genomes of nine *Phyllanthus* species display a unique IR contraction and expansion type.

### 3.5 Genome comparison and nucleotide diversity

A comparison of overall sequence variation showed that the cp genome of *Phyllanthus* is quite different. The sequence divergence of IR regions was lower than that of SSC and LSC regions, and the coding region was more conserved than the non-coding regions. Except for the more remarkable mutations in the *ndhF*, *ycf1*, and *ycf2* genes, all protein-coding genes showed a highly conserved character. The highest divergence was mainly found in intergenic

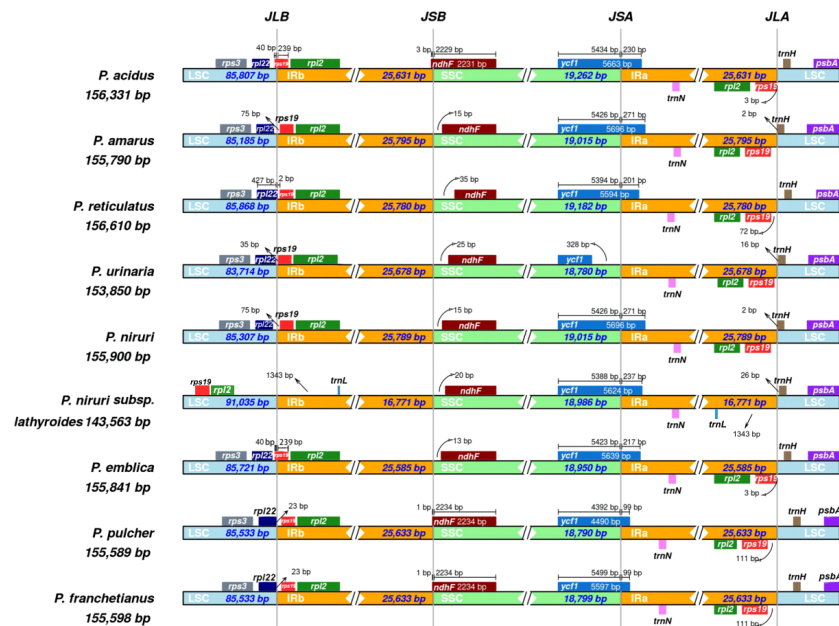


FIGURE 3

Comparisons of the borders of LSC, SSC, and IRa/b regions among the nine *Phyllanthus* cp genomes. The numbers represent the distance between the gene ends and the border sites, and the numbers below represent the length of the LSC, SSC, and IRa/b regions.

spacers (IGS), such as *rps16-trnQ-UUG*, *trnS-GCU-trnG-UCC*, *trnG-UCC-trnR-UCU*, *trnE-UUC-trnT-GGU*, *trnD-GUC-trnY-GUA*, *trnT-UGU-trnL-UAA*, *trnL-CAA-ndhB*, *trnN-GUU-trnR-ACG*, and *rps15-ycf1* (Figure 4). Besides, the sliding window analysis demonstrated that the nine regions, including *rps16*, *trnS-GCU-trnG-UCC*, *trnG-UCC-trnR-UCU*, *petA-psbJ*, *rps3*, *rrn5S-trnR-ACG*, *ndhF*, *ndhE-ndhG*, and *ycf1*, had higher nucleotide diversity values ( $Pi > 0.05$ ) (Figure S3). The results above show that 12 highly variable sites (*rps16-trnQ-UUG*, *trnS-GCU-trnG-UCC*, *trnG-UCC-trnR-UCU*, *trnE-UUC-trnT-GGU*, *trnD-GUC-trnY-GUA*, *trnT-UGU-trnL-UAA*, *trnL-CAA-ndhB*, *trnN-GUU-trnR-ACG*, *rps15-ycf1*, *petA-psbJ*, *rrn5S-trnR-ACG*, and *ndhE-ndhG*) might be able to be used as molecular markers to identify *P. urinaria* and its contaminants.

### 3.6 Species authentication analysis based on IGS

IGS regions are the most commonly used markers for phylogenetic studies at plant taxonomic levels, as they are regarded as more variable and may provide more phylogenetically informative characters. To find candidate sequences for identifying *P. urinaria* and its adulterants, 12 IGS were extracted from 13 *Phyllanthus* species using PhyloSuite v1.2.2. And each of them is subject to ML analyses in IQtree. As shown in Figure S4.1-4.12, *P. urinaria* could be distinguished from its

common adulterants based on *trnS-GCU-trnG-UCC*, *trnT-UGU-trnL-UAA*, and *petA-psbJ*, whereas the remaining IGS cannot be distinguished, and the bootstrap values for the relationship among these clades were weak (<70%). Furthermore, the ML phylogenetic tree was also inferred using a combination of these three IGS. The results (Figure S5) showed that *P. urinaria* (Genbank OP009346) was located in independent branches and that there was a well-supported sister relationship between *P. urinaria* (Genbank OP009346) and *P. amarus* (Genbank OP009344) + *P. urinaria* (Genbank NC060522) + *P. niruri* (Genbank OP009347). These results indicated that combining three IGS could effectively discriminate *P. urinaria* from its common adulterants.

### 3.7 Phylogenetic analysis and divergence time analysis

The ML and NJ phylogenetic trees were inferred using 40 species, with two *Daphniphyllum* species as outgroups. The consensus trees obtained from the inference analyses showed that most nodes resolved with high support (Figure 5, Figure S6). The phylogenetic trees generated by the ML and NJ alignments have similar topologies. Each subfamily of the Euphorbiaceae family forms a monophyletic clade. Acalyphoideae and Euphorbioideae + Crotonoideae were sister taxa within the four subfamilies, and Phyllanthoideae was a sister group to the clade containing Acalyphoideae + Crotonoideae +

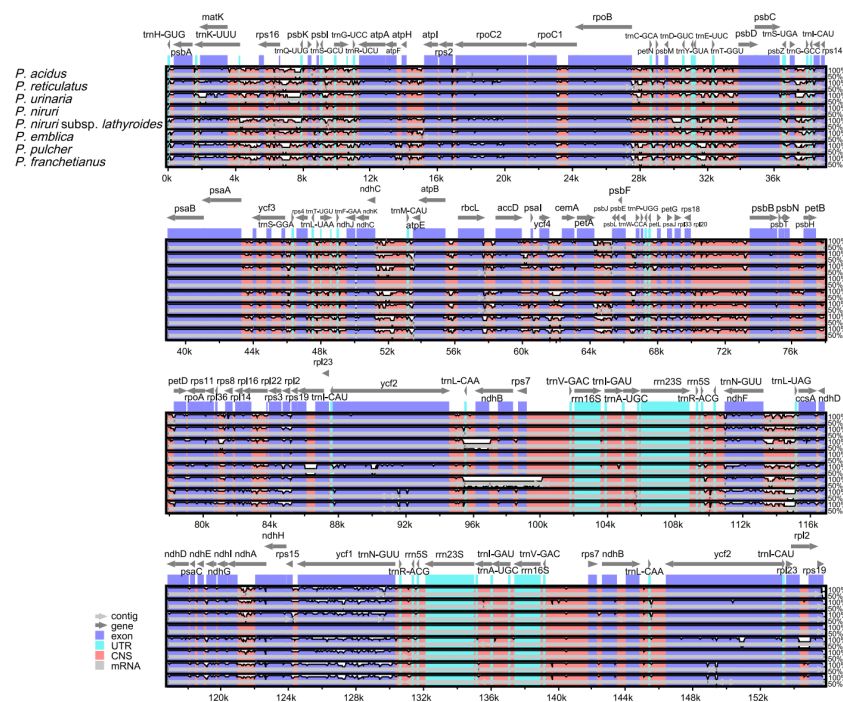


FIGURE 4

Comparison of nine cp genomes using *P. amarus* annotation as a reference. The vertical scale indicates the percentage of identity, ranging from 50% to 100%. The horizontal axis shows the coordinates within the cp genome. Genome regions are color-coded as exons, introns, and IGS, and the gray arrows indicate the direction of transcription of each gene. Annotated genes are displayed along the top.

Euphorbioideae. In addition, the subfamily Phyllanthoideae is further divided into three clades: (i) clade A, including *Glochidion wrightii* (Genbank MW801302), *G. eriocarpum* (Genbank MW801303), *G. chodoense* (Genbank NC042906), *G. puberum* (Genbank MW801304), *P. amarus* (Genbank NC047474), *P. emblica* (Genbank MN122078, Genbank OP009349, Genbank NC047477), *P. acidus* (Genbank OP009343), and *P. niruri* subsp. *lathyroides* (Genbank OP009348); (ii) clade B, including *P. amarus* (Genbank OP009344), *P. urinaria* (Genbank NC060522), *P. niruri* (Genbank OP009347), and *P. urinaria* (Genbank OP009346); (iii) clade C, including *P. franchetianus* (Genbank OP009351), *P. pulcher* (Genbank OP009350), and *P. reticulatus* (Genbank OP009345).

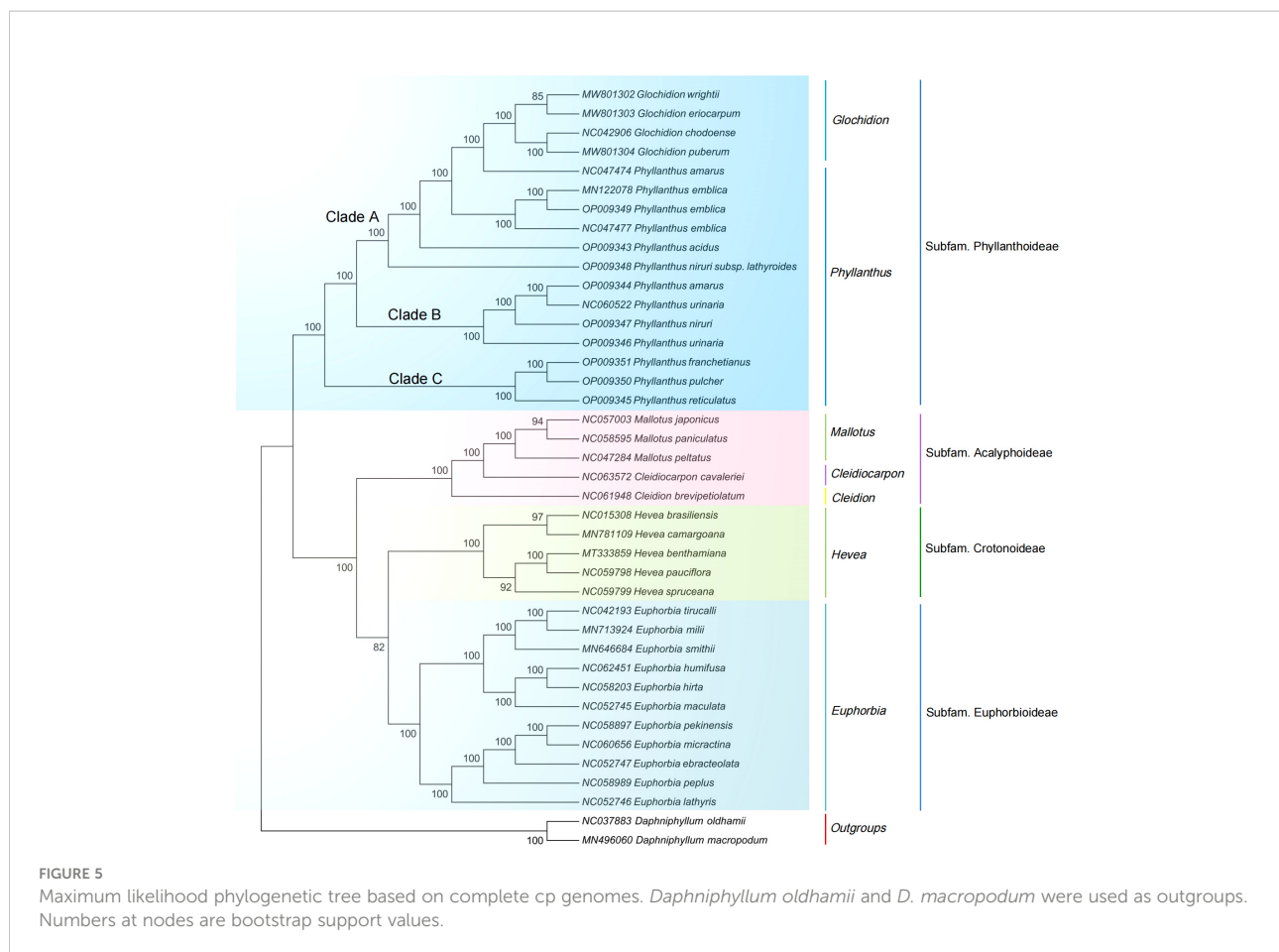
In addition, 34 cp genomes of Euphorbiaceae family plants (including 13 *Phyllanthus* species) and two outgroups are used to estimate the divergence times. *Phyllanthus* were estimated to have originated 48.72 million years ago (Ma). The two main lineages, clade A + clade B and clade C, seem to have radiated since the Oligocene (clade A + clade B: 36.85 Ma; clade C: 30.89 Ma; Figure 6). The extant subfamilies of the Euphorbioideae and Acalyphoideae shared a common ancestor at the beginning of the Cretaceous (91.91 Ma), while the split between the Euphorbioideae and Crotonoideae is estimated to have occurred at 75.64 Ma. Moreover, the divergence times of

the *Daphniphyllum* occurred at 120.77 Ma, having shared a common ancestor with the Euphorbiaceae.

## 4 Discussion

### 4.1 Cp genome structure and comparative analysis

In the present study, the cp genomes of nine species of *Phyllanthus* were assembled *de novo* and analyzed comparatively. The observed genome size is within the size range known for most angiosperms, ranging from 107 kb in *Cathaya argyrophylla* (Pinaceae) to 218 kb in *Pelargonium* (geraniums; Geraniaceae) (Daniell et al., 2016). Besides, our findings revealed that the total length, GC content, and gene composition of the cp genomes were almost identical in all species. Previous studies have shown that the cp genome of angiosperms is highly conserved at the genus level (Khan et al., 2020; Villanueva-Corralles et al., 2021; Feng et al., 2022). Moreover, previous research demonstrated that the *rps19* gene existed in the IR region (Ahmed et al., 2012; Wang et al., 2022). In contrast, the *rps19* of *P. niruri* subsp. *lathyroides* is located entirely in the LSC Region, possibly due to IR contraction. The same pattern was reported in *Anchomanes hookeri*, and *Peucedanum* (Henriquez et al., 2020; Liu et al., 2022).



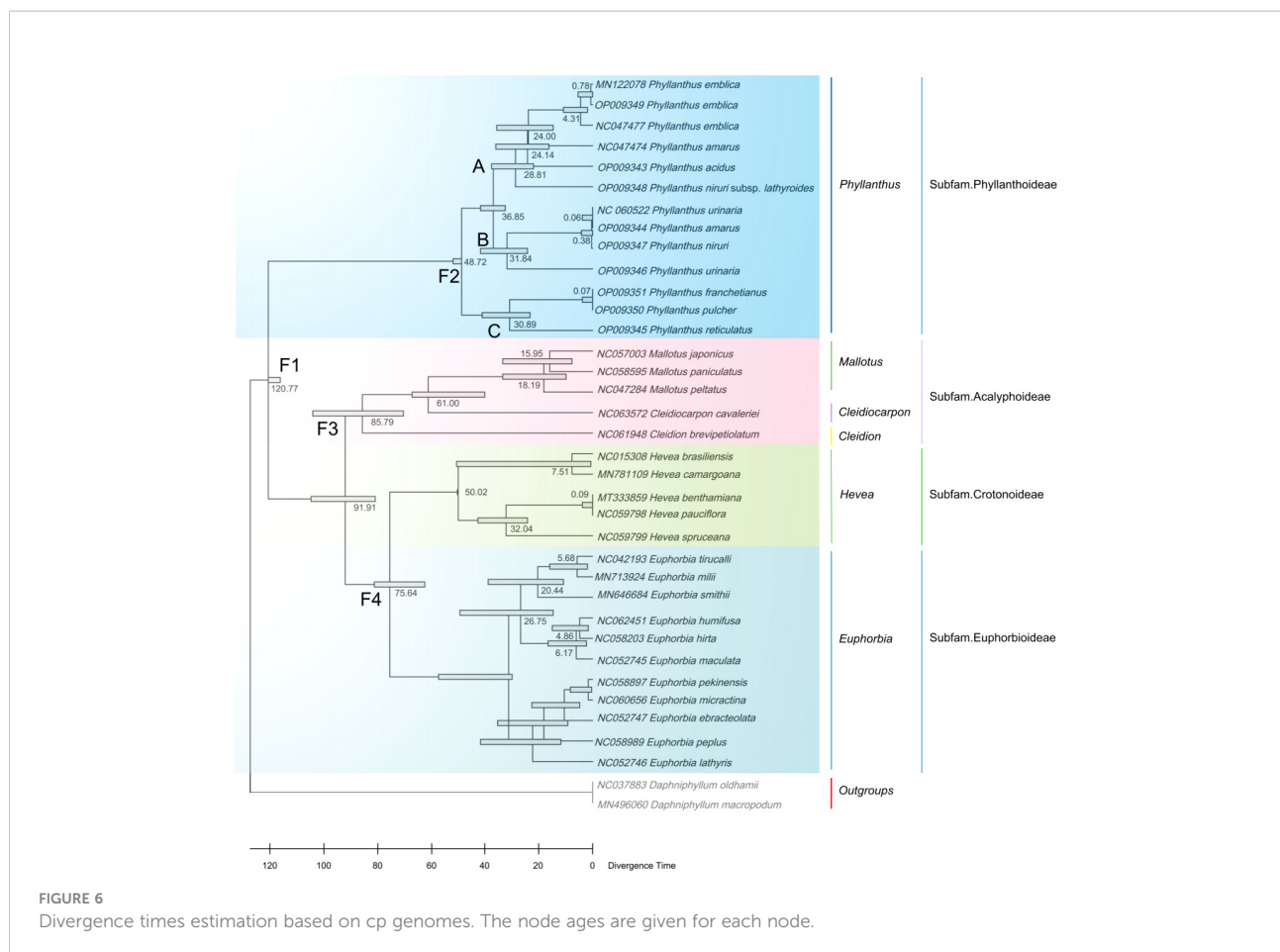
## 4.2 Species identification and phylogenetic analysis

Previous molecular studies of the *Rheum*, *Hedyotis*, and *Curcuma* species showed that cp genetic markers had high identification capabilities (Zhou et al., 2018; Gui et al., 2020; Yik et al., 2021). Our study revealed that three regions (*trnS-GCU-trnG-UCC*, *trnT-UGU-trnL-UAA*, and *petA-psbJ*) might be potential molecular markers for identifying *P. urinaria* and its common adulterants. Bouman et al. (2021) found that the *trnS-GCU-trnG-UCC* could distinguish *Phyllanthus* species. Notably, Zhang et al. (2021) also found that the *trnS-GCU-trnG-UCC* could be potential molecular markers for distinguishing *Alpinia* species. Moreover, *trnT-UGU-trnL-UAA* or *petA-psbJ* were reported as potential markers for other species identification (Dong et al., 2021; Wu et al., 2021). Although these previous studies revealed that universal DNA barcode (e.g., *psbA-trnH*) could differentiate *P. urinaria* from their related species (Srirama et al., 2010; Inglis et al., 2018), some common adulterants were not included in these studies. Furthermore, the comparative analysis showed that the screened IGS exhibited higher variability than *psbA-trnH*. Theoretically, these IGS could differentiate nine selected species, whereas a much

more thorough investigation of identification accuracy and amplification efficiency is required, as well as more experimental evidence.

Moreover, ML analysis demonstrated that *P. urinaria* (Genbank OP009346) was located in independent branches in the phylogeny and strongly supported the sister relationship between *P. urinaria* (Genbank OP009346) and the well-supported clade (*P. amarum*, Genbank OP009344; and *P. niruri*, Genbank OP009347). The results indicated that the cp genome could discriminate *P. urinaria* from *P. amarum* and *P. niruri*, which was supported by the findings of other researchers based on ITS, *matK*, *psbA-trnH*, *trnL*, and *trnL-trnF* (Inglis et al., 2018). However, the samples of *P. urinaria* (Genbank OP009346) and *P. urinaria* (Genbank NC060522) were not recovered as monophyletic and were placed in different branches. In the previous study, some researchers found that intraspecific diversity existed in *Isodon rubescens* and *Artemisia argyi* from different geographical areas (Zhou et al., 2022; Chen et al., 2022). Therefore, the difference in geographical origins may explain why the two species are split in these clades. Besides, both NJ and ML analyses found strong support for a sister relationship between *P. reticulatus* (Genbank OP009345) and *P. pulcher* (Genbank OP009350), which agreed with the findings of Hidalgo et al. (2020) based on *trnK-matK*, *matK*, ITS, and *matK*+ITS





(Pagare et al., 2016). In a previous study, Rehman et al. (2021) also found that the polymorphic protein-coding genes, including *rpl22*, *ycf1*, *matK*, *ndhF*, and *rps15*, may help reconstruct the high-resolution phylogenetic tree of the family Phyllanthaceae. In general, our results provide a valuable reference and a foundation for using cp genomes in species identification and aid in the understanding of the phylogeny of *Phyllanthus*.

### 4.3 Divergence time of *Phyllanthus*

According to divergence time estimates, the early divergence of *Phyllanthus* occurred at approximately 48.72 Ma during the early Eocene, which is congruent with other studies (Kawakita and Kato, 2009; Welzen et al., 2015). Since the late Eocene, the previous study reported that the global climate started to have a notable change; as the humidity and precipitation gradually increased (Zachos et al., 2001) and slowly cooled within this timeframe. These climate changes may have promoted the dispersals/migrations, and diversification of land plants (Zuo et al., 2017). In addition, Dynesius and Jansson (2000) and Zachos et al. (2001) also found that the temperature increase affected various plant and animal communities and groups at the

Oligocene/Miocene boundary (~23 Ma). Among the effects of climate change, there was likely increased speciation in the niches that opened after the end of the climatic fluctuations (Bouman et al., 2021). Therefore, we can conclude that the climatic changes may have contributed to the diversification of *Phyllanthus* during the Eocene.

## 5 Conclusion

In the present study, the complete cp genomes of nine species of *Phyllanthus* were *de novo* assembled from high throughput sequencing reads, and the cp genomes of *P. acidus*, *P. reticulatus*, *P. niruri*, *P. niruri* subsp. *lathyroides*, *P. pulcher*, and *P. franchetianus* were reported for the first time. These cp genomes were generally conserved and exhibited similar gene content and genomic structure. Three highly variable cp loci, including *trnS-GCU-trnG-UCC*, *trnT-UGU-trnL-UAA*, and *petA-psbJ* were identified and could serve as candidate markers for identifying *P. urinaria* and its common adulterants. Meanwhile, the complete cp genome was considered a reliable molecular marker for identifying these species, which may have virtual significance for protecting their diversity and making management decisions

for this species. The divergence of *Phyllanthus* from ancestral taxa occurred in the early Eocene, which might be due to geological and climatic changes. In conclusion, our study provides a powerful tool and valuable scientific reference for the safety and effectiveness of clinical drug use, and it also contributes to the bioprospecting and conservation of *Phyllanthus* species.

## Data availability statement

The data presented in the study are deposited in the GenBank repository, accession numbers were from OP009343 to OP009351.

## Author contributions

HF, YL, and PZ participated in the conception and design of the research. HF, GD and BL collected the species. HF and GD are responsible for analyzing and processing data. HF wrote the manuscript. The paper was revised by YL and PZ. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Yunnan academician expert workstation (202105AF150053), the key technology projects in the Yunnan province of China (202002AA100007).

## References

- Abdullah, M., Mehmood, F., Shahzadi, I., Waseem, S., Mirza, B., Ahmed, I., et al. (2020). Chloroplast genome of hibiscus rosa-sinensis (Malvaceae): Comparative analyses and identification of mutational hotspots. *Genomics* 112, 581–591. doi: 10.1016/j.ygeno.2019.04.010
- Adedapo, A. A., Abatan, M. O., Idowu, S. O., and Olorunsogo, O. O. (2005). Toxic effects of chromatographic fractions of *Phyllanthus amarus* on the serum biochemistry of rats. *Phytother. Res.* 19, 812–815. doi: 10.1002/ptr.1721
- Ahmed, I., Biggs, P. J., Matthews, P. J., Collins, L. J., Hendy, M. D., and Lockhart, P. J. (2012). Mutational dynamics of aroid chloroplast genomes. *Genome Biol. Evol.* 4, 1316–1323. doi: 10.1093/gbe/evs110
- Allen, G. C., Flores-Vergara, M. A., Krasynanski, S., Kumar, S., and Thompson, W. F. (2006). A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* 1, 2320–2325. doi: 10.1038/nprot.2006.384
- Amiryousefi, A., Hyvönen, J., and Pocza, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34, 3030–3031. doi: 10.1093/bioinformatics/bty220
- Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics* 33, 2583–2585. doi: 10.1093/bioinformatics/btx198
- Bouman, R. W., Kessler, P. J. A., Telford, I. R. H., Bruhl, J. J., Strijk, J. S., Saunders, R. M. K., et al. (2021). Molecular phylogenetics of *Phyllanthus sensu lato* (Phyllanthaceae): towards coherent monophyletic taxa. *Taxon* 70, 72–98. doi: 10.1002/tax.12424
- Brudno, M., Malde, S., Poliakov, A., Do, C. B., Couronne, O., Dubchak, L., et al. (2003). Glocal alignment: Finding rearrangements during alignment. *Bioinformatics* 19, i54–i62. doi: 10.1093/bioinformatics/btg1005
- Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., et al. (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* 181, 1–20. doi: 10.1111/boj.12385
- Chen, C. J., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y. H., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, C. J., Miao, Y. H., Luo, D. D., Li, J. X., Wang, Z. X., Luo, M., et al. (2022). Sequence characteristics and phylogenetic analysis of the *Artemisia argyi* chloroplast genome. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.906725
- Chudapongse, N., Kamkhunthod, M., and Poompachee, K. (2010). Effects of *Phyllanthus urinaria* extract on HepG2 cell viability and oxidative phosphorylation by isolated rat liver mitochondria. *J. Ethnopharmacol.* 130, 315–319. doi: 10.1016/j.jep.2010.05.010
- Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Bio.* 17, 134. doi: 10.1186/s13059-016-1004-2
- Dong, W. P., Liu, Y. L., Xu, C., Gao, Y. W., Yuan, Q. J., Suo, Z. L., et al. (2021). Chloroplast phylogenomic insights into the evolution of *Distylium* (Hamamelidaceae). *BMC Genomics* 22, 293. doi: 10.1186/s12864-021-07590-6
- Dynesius, M., and Jansson, R. (2000). Evolutionary consequences of changes in species' geographical distributions driven by milankovitch climate oscillations. *P. Natl. Acad. Sci. U.S.A.* 97, 9115–9120. doi: 10.1073/pnas.97.16.9115
- Feng, Y., Gao, X. F., Zhang, J. Y., Jiang, L. S., Li, X., Deng, H. N., et al. (2022). Complete chloroplast genomes provide insights into evolution and phylogeny of *Campylotropis* (Fabaceae). *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.895543

## Acknowledgments

We would like to thank Mr. Haitao Li for his assistance in obtaining specimens for this study. We also thank Yuan Jiang helped to teach the software used for the experiments.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1099856/full#supplementary-material>

- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 32, W273–W279. doi: 10.1093/nar/gkh458
- Geethangili, M., and Ding, S. (2018). A review of the phytochemistry and pharmacology of *Phyllanthus urinaria* L. *Front. Pharmacol.* 9. doi: 10.3389/fphar.2018.01109
- Geng, H. C., Zhu, H. T., Yang, W. N., Wang, D., Yang, C. R., and Zhang, Y. J. (2021). New cytotoxic dichapetalins in the leaves of *Phyllanthus acidus*: Identification, quantitative analysis, and preliminary toxicity assessment. *Bioorg. Chem.* 114, 105125. doi: 10.1016/j.bioorg.2021.105125
- Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* 47, W59–W64. doi: 10.1093/nar/gkz238
- Guan, Y. H., Liu, W. W., Duan, B. Z., Zhang, H. Z., Chen, X. B., Wang, Y., et al. (2022). The first complete chloroplast genome of *Vicatia thibetica* de Boiss.: Genome features, comparative analysis, and phylogenetic relationships. *Physiol. Mol. Biol. Plants* 28, 439–454. doi: 10.1007/s12298-022-01154-y
- Gui, L. J., Jiang, S. F., Xie, D. F., Yu, L. Y., Huang, Y., Zhang, Z. J., et al. (2020). Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. *Gene* 732, 144355. doi: 10.1016/j.gene.2020.144355
- Guo, Q., Zhang, Q. Q., Chen, J. Q., Zhang, W., Qiu, H. C., Zhang, Z. J., et al. (2017). Liver metabolomics study reveals protective function of *Phyllanthus urinaria* against CCl<sub>4</sub>-induced liver injury. *Chin. J. Nat. Medicines* 15, 525–533. doi: 10.1016/S1875-5364(17)30078-X
- Heinze, B. (2007). A database of PCR primers for the chloroplast genomes of higher plants. *Plant Methods* 3, 4. doi: 10.1186/1746-4811-3-4
- Henriquez, C. L., Abdullah, A., Ahmed, I., Carlsen, M. M., Zuluaga, A., Croat, T. B., et al. (2020). Evolutionary dynamics of chloroplast genomes in subfamily *Aroideae* (Araceae). *Genomics* 112, 2349–2360. doi: 10.1016/j.ygeno.2020.01.006
- Hidalgo, B. F., Bazan, S. F., Iturralde, R. B., and Borsch, T. (2020). Phylogenetic relationships and character evolution in neotropical *Phyllanthus* (Phyllanthaceae), with a focus on the Cuban and Caribbean taxa. *Int. J. Plant Sci.* 181, 284–305. doi: 10.1086/706454
- Inglis, P., Mata, L., Da Silva, M., Vieira, R., Alves, R. D. B. N., Silva, D., et al. (2018). DNA Barcoding for the identification of *Phyllanthus* taxa used medicinally in Brazil. *Planta Med.* 84, 1300–1310. doi: 10.1055/a-0644-2688
- Jiang, Y., Miao, Y. J., Qian, J., Zheng, Y., Xia, C. L., Yang, Q. S., et al. (2022). Comparative analysis of complete chloroplast genome sequences of five endangered species and new insights into phylogenetic relationships of *Paris*. *Gene* 833, 146572. doi: 10.1016/j.gene.2022.146572
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biol.* 21, 241–272. doi: 10.1186/s13059-020-02154-5
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kawakita, A., and Kato, M. (2009). Repeated independent evolution of obligate pollination mutualism in the phyllanthaceae-epicephala association. *P Roy Soc. B-Biol Sci.* 276, 417–426. doi: 10.1098/rspb.2008.1226
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199
- Khan, A., Asaf, S., Khan, A. L., Shehzad, T., Al-Rawahi, A., and Al-Harrasi, A. (2020). Comparative chloroplast genomics of endangered euphorbia species: Insights into hotspot divergence, repetitive sequence variation, and phylogeny. *Plants* 9, 199. doi: 10.3390/plants9020199
- Kiran, K. R., Swathy, P. S., Paul, B., Shama Prasad, K., Radhakrishna Rao, M., Joshi, M. B., et al. (2021). Untargeted metabolomics and DNA barcoding for discrimination of *Phyllanthus* species. *J. Ethnopharmacol.* 273, 113928. doi: 10.1016/j.jep.2021.113928
- Kumar, S., Stecher, G., Li, M., Niyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29, 4633–4642. doi: 10.1093/nar/29.22.4633
- Li, F., Liu, Y., Wang, J. H., Xin, P. Y., Zhang, J. Y., Zhao, K., et al. (2022). Comparative analysis of chloroplast genome structure and phylogenetic relationships among six taxa within the genus *Catalpa* (Bignoniaceae). *Front. Genet.* 13. doi: 10.3389/fgene.2022.845619
- Li, X. W., Yang, Y., Henry, R. J., Rossetto, M., Wang, Y. T., and Chen, S. L. (2015). Plant DNA barcoding: From gene to genome. *Biol. Rev.* 90, 157–166. doi: 10.1111/brv.12104
- Liu, C. K., Lei, J. Q., Jiang, Q. P., Zhou, S. D., and He, X. J. (2022). The complete plastomes of seven *Peucedanum* plants: comparative and phylogenetic analyses for the peucedanum genus. *BMC Plant Biol.* 22, 101. doi: 10.1186/s12870-022-03488-x
- Manissorn, J., Sukrong, S., Ruangrunsi, N., and Mizukami, H. (2010). Molecular phylogenetic analysis of *Phyllanthus* species in Thailand and the application of polymerase chain reaction-restriction fragment length polymorphism for *Phyllanthus amarus* identification. *Biol. Pharm. Bull.* 33, 1723–1727. doi: 10.1248/bpb.33.1723
- Mello, B. (2018). Estimating TimeTrees with MEGA and the TimeTree resource. *Mol. Biol. Evol.* 35, 2334–2342. doi: 10.1093/molbev/msy133
- Neuhaus, H. E., and Emes, M. J. (2000). Nonphotosynthetic metabolism in plastids. *Annu. Rev. Plant Biol.* 51, 111–140. doi: 10.1146/annurev.arplant.51.1.111
- Nguyen, L., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-Tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Pagare, R., Naik, S., Krishnan, S., and Janarthanam, M. (2016). Lectotypification of pseudoglochidion anomalayanum gamble and its taxonomic position under the genus *Phyllanthus* (Phyllanthaceae). *Phytotaxa* 286, 61–74. doi: 10.11646/phytotaxa.286.2.1
- Pruesapan, K., Telford, I. R. H., Bruhl, J. J., and van Welzen, P. C. (2012). Phylogeny and proposed circumscription of *Breynia*, *Sauropus* and *Symostemon* (Phyllanthaceae), based on chloroplast and nuclear DNA sequences. *Aust. Syst. Bot.* 25, 313. doi: 10.1071/SB11005
- Rehman, U., Sultana, N., Abdullah, A., Jamal, A., Muzaffar, M., and Poczi, P. (2021). Comparative chloroplast genomics in *Phyllanthaceae* species. *Diversity (Basel)* 13, 403. doi: 10.3390/d13090403
- Ren, J., Tian, J., Jiang, H., Zhu, X. X., Mutie, F. M., Wanga, V. O., et al. (2022). Comparative and phylogenetic analysis based on the chloroplast genome of *Coleanthus subtilis* (Tratt.) Seidel, a protected rare species of monotypic genus. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.828467
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Sarin, B., Clemente, J. P. M., and Mohanty, A. (2013). PCR-RFLP to distinguish three *Phyllanthus* sp., commonly used in herbal medicines. *S. Afr. J. Bot.* 88, 455–458. doi: 10.1016/j.sajb.2013.09.011
- Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689. doi: 10.1093/nar/gki366
- Sharp, P. M., Tuohy, T. M., and Mousurki, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14, 5125–5143. doi: 10.1093/nar/14.13.5125
- Shi, L. C., Chen, H. M., Jiang, M., Wang, L. Q., Wu, X., Huang, L. F., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65–W73. doi: 10.1093/nar/gkz345
- Srirama, R., Senthikumar, U., Sreejayan, N., Ravikanth, G., Gurumurthy, B. R., Shivanna, M. B., et al. (2010). Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools. *J. Ethnopharmacol.* 130, 208–215. doi: 10.1016/j.jep.2010.04.042
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45, W6–W11. doi: 10.1093/nar/gkx391
- Villanueva-Corralles, S., García-Botero, C., Garcés-Cardona, F., Ramírez-Rios, V., Villanueva-Mejía, D. F., and Álvarez, J. C. (2021). The complete chloroplast genome of *Plukenetia volubilis* provides insights into the organelle inheritance. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.667060
- Wang, J., Qian, J., Jiang, Y., Chen, X. C., Zheng, B. J., Chen, S. L., et al. (2022). Comparative analysis of chloroplast genome and new insights into phylogenetic relationships of *Polygonatum* and tribe *Polygonateae*. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.882189
- Webster, G. L., and Carpenter, K. J. (2008). Pollen morphology and systematics of palaeotropical *Phyllanthus* and related genera of subtribe *Phyllanthinae* (Euphorbiaceae). *Bot. J. Linn. Soc.* 157, 591–608. doi: 10.1111/j.1095-8339.2008.00781.x
- Welzen, P. C., Pruesapan, K., Telford, I. R. H., and Bruhl, J. J. (2015). Historical biogeography of *Breynia* (Phyllanthaceae); what caused speciation? *J. Biogeogr.* 42, 1493–1502. doi: 10.1111/jbi.12517

- Wu, L. W., Cui, Y. X., Wang, Q., Xu, Z. C., Wang, Y., Lin, Y. L., et al. (2021). Identification and phylogenetic analysis of five *Crataegus* species (Rosaceae) based on complete chloroplast genomes. *Planta* 254, 14. doi: 10.1007/s00425-021-03667-4
- Xiong, C., Sun, W., Li, J. J., Yao, H., Shi, Y. H., Wang, P., et al. (2018). Identifying the species of seeds in traditional Chinese medicine using DNA barcoding. *Front. Pharmacol.* 9. doi: 10.3389/fphar.2018.00701
- Yik, M. H., Kong, B. L., Siu, T., Lau, D. T., Cao, H., and Shaw, P. (2021). Differentiation of *Hedyotis diffusa* and common adulterants based on chloroplast genome sequencing and DNA barcoding markers. *Plants* 10, 161. doi: 10.3390/plants10010161
- Zachos, J., Pagani, M., Sloan, L., Thomas, E., and Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292, 686–693. doi: 10.1126/science.1059412
- Zhang, Y., Song, M. F., Li, Y., Sun, H. F., Tang, D. Y., Xu, A. S., et al. (2021). Complete chloroplast genome analysis of two important medicinal *Alpinia* species: *Alpinia galanga* and *Alpinia kwangsiensis*. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.705892
- Zhou, Y. X., Nie, J., Xiao, L., Hu, Z. G., and Wang, B. (2018). Comparative chloroplast genome analysis of rhubarb botanical origins and the development of specific identification markers. *Molecules* 23, 2811. doi: 10.3390/molecules23112811
- Zhou, Z. Y., Wang, J., Pu, T. T., Dong, J. J., Guan, Q., Qian, J., et al. (2022). Comparative analysis of medicinal plant *Isodon rubescens* and its common adulterants based on chloroplast genome sequencing. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.1036277
- Zuo, Y. J., Wen, J., and Zhou, S. L. (2017). Intercontinental and intracontinental biogeography of the eastern Asian-Eastern north American disjunct *Panax* (the ginseng genus, *Araliaceae*), emphasizing its diversification processes in eastern Asia. *Mol. Phylogenet. Evol.* 117, 60–74. doi: 10.1016/j.ympev.2017.06.016