

Adaptive Evolution and Phylogenetic Analyses in *Pinellia* based on Chloroplast Genomes

Ning Cui (✉ cuning20061988@126.com)

Shandong Academy of Chinese Medicine <https://orcid.org/0000-0002-5478-9743>

Weixu Chen

Shang yao hua yu (LinYi) Traditional Chinese Medicine Resources Co., Ltd.

Xiwen Li

China Academy of Chinese Medical Sciences Institute of Chinese Materia Medica

Ping Wang

Shandong Academy of Chinese Medicine

Research Article

Keywords: *Pinellia*, Phylogeny, Evolution, Divergence time estimation, Chloroplast genome

Posted Date: October 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-985139/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background *Pinellia* Tenore (Araceae) is a genus of perennial herbaceous plants, all of which have medicinal value. The chloroplast (cp) genome data of *Pinellia* are scarce, and the phylogenetic relationship and gene evolution remain unclear.

Methods and results In this study, we sequenced and annotated the *Pinellia pedatisecta* cp genome and combined it with previously published genomes for other *Pinellia* species. We used bioinformatics methods to analyse the genomic structure, repetitive sequences, interspecific variation, divergence hotspots, phylogenetic relationships, divergence time estimation and selective pressure of four *Pinellia* plastomes. Results showed that the cp genomes of *Pinellia* varied in length between 168,178 (*P. pedatisecta* MN046890) and 164,013 bp (*P. ternata* KR270823). A total of 68–111 SSR loci were identified as candidate molecular markers for further genetic diversity study. Eight mutational hotspot regions were determined, including *psbI-trnG-UCC*, *psbM-rpoB*, *ndhJ-trnT-UGU*, *trnP-UGG-trnW-CCA*, *ndhF-trnN-GUU*, *ndhG-ndhE*, *ycf1-rps15* and *trnR-ycf1*. Gene selection pressure suggested that four genes were subjected to positive selection. Phylogenetic inferences based on the complete cp genomes and their protein-coding sequences revealed a sister relationship between *Pinellia* and *Arisaema* plants whose divergence was estimated to occur around 22.48 million years ago. All *Pinellia* plants formed a monophyletic evolutionary clade in which *P. peltata*, rather than *P. pedatisecta*, earlier diverged, indicating that *P. pedatisecta* is not the basal taxon of *Pinellia* but *P. peltata* may be.

Conclusions The cp genomes of *Pinellia* will provide valuable information for species classification, identification, molecular breeding and evolutionary exploration of the genus *Pinellia*.

Introduction

Pinellia Tenore is a small eastern Asian genus in the Araceae family. Although there are only seven perennial herbaceous species in *Pinellia* genus [1], every member it contains is important traditional Chinese medicinal plant and was recorded in Chinese herb classics more than 2000 years ago; the most famous among them is *P. ternata*, with an annual demand of 5,500–6,000 tons [2]. They have been used for the treatment of viper bites, lumbago, allergic reaction and externally to treat traumatic injury, abscesses, neck lymphosarcoma, breast mastitis and uterine cancer [3, 4].

In recent years, the phylogeny and evolution of monocots have come under intense scrutiny with the rapid development of molecular phylogenetic systematics, and multiple studies have highlighted the phylogeny of *Pinellia* as problematic [5, 6]. The phylogenetic position of the *Pinellia* genus in Araceae has been controversial, and sister groups of *Pinellia* show discrepancy under different classification systems and studies [7, 8]. The six major taxonomic systems of Araceae, namely, Schott system [9], Engler system [10], Hutchinson system [11], Grayum system [12], Bogner and Nicolson system [13] and Mayo et al. system [14], have different views on the sister group of *Pinellia*, with four genera (*Cryptocoryne*, *Langenandra*, *Ambrosina* and *Arisaema*) as candidates. With the development of molecular systematics, analyses of the plastome gene and restriction-site sequences suggested that the *Arisaema* genus is strongly related to *Pinellia* and is its sister group [7, 15, 16]. Contrary to the above classification, Keating [17] and Bogner and Petersen [18] collected morpho-anatomical data and argued that *Arisaema* and *Pinellia* cannot gather into a unique clade for their morphological discrepancy of stamens.

Members of *Pinellia* exhibit wide variations in flower, leaf, bulblet, spathe and ovule characteristics [19, 20]. The low-level taxonomy and interspecific phylogeny of *Pinellia* are difficult to address based on morphological traits. The most comprehensive below-genus phylogenetic analysis to date has been provided by Yin, a study using a matrix of all seven *Pinellia* species and ITS and *trnL-F* DNA barcode sequences [21]. Combined with the morphological traits, Yin concluded that *P. pedatisecta* is the basal taxon of *Pinellia* and suggested that its taxonomic rank should be elevated to a section. However, our group previously conducted interspecific phylogenetic analyses of *Pinellia* by using four DNA barcode sequences (*ITS*, *matK*, *rcbL* and *trnL-F*) with the same method as Yin. The topologies of four barcodes were not consistent, with *P. pedatisecta* in the outer layer of ITS and *matK* trees and in the inner layer of *rcbL* and *trnL-F* trees, receiving considerably lower bootstrap values (Supplementary Fig. S1) probably due to insufficient sequence length and interspecific variations.

The chloroplast (cp) is an important self-replicating organelle that plays a crucial role in photosynthesis and in the synthesis of pigment, protein and starch [22]. The cp contains its own circular double-stranded genome, which is inherited maternally in most angiosperms or paternally in some gymnosperms [23]. Unlike the nuclear genome, the cp lacks meiotic recombination. These properties, along with adequate levels of polymorphism, make it a suitable molecule for studies on phylogeny and evolution [24]. Although some researchers have used complete cp genome sequences to explore the phylogenetic relationships of some species in *Pinellia* [22, 23, 25, 26], which revealed that *Pinellia* species were clustered into a group [22, 25], and *P. peltata* diverged relatively earlier [25], these studies ignored the evolution of whole cp genomes of *Pinellia*, including, for example, the estimation of genes selection pressure and divergence time.

To reveal the interspecific diversification pattern of *Pinellia*, we determined to sequence the cp genome of *P. pedatisecta* by high-throughput sequencing after personal identification and sampling. Here, we determined the complete cp genome of *P. pedatisecta* and compared its sequence features with three other *Pinellia* plastomes. The main goals of this study were to (1) characterize and compare the cp genomes of *Pinellia* and detect the differences between *Pinellia* species and between published and newly assembled cp genomes of *P. pedatisecta*; (2) identify simple sequence repeats (SSRs), long repeats and genetically variable regions and select eight divergence hotspots as candidate DNA barcodes; (3) reconstruct phylogenetic relationships of *Pinellia* species based on the cp genome alignments and verify their phylogenetic position within Araceae;

and (4) estimate genes selection pressure and divergence time for determining the relative order and spacing of speciation events of *Pinellia*. Comparative cp genomic analysis could provide theoretical basis to further understand the evolution of the Araceae family and additional insights into the long-standing controversial intergenus and intragenus phylogeny of *Pinellia*.

Materials And Methods

Plant material, DNA extraction, sequencing, assembly and annotation

In our trial, the cp genome of *P. pedatisecta* was sequenced to explore the phylogeny and evolution of *Pinellia*. Fresh leaves of *P. pedatisecta* from Linyi City, Shandong Province were sampled. Voucher specimens were deposited in Shandong Academy of Chinese Medicine. Total genomic DNA was extracted from 100 mg of silica-dried leaf by using a DNeasy Plant MiniKit (Qiagen, CA, USA) according to the manufacturer's instructions [27]. The quantity and quality of genomic DNA were examined by using ND-2000 spectrometer (ThermoFisher Scientific, Wilmington, DE, USA) and 0.8% agarose gel electrophoresis. The chloroplast genome of *Salvia plebeia* (NC050929) was assembled in our previous study [27]. Taking this work as a guidance, the DNA sample pre-treatment, whole genome sequencing, cp genome assembly, junction validation, and annotation were performed in turn with the cp genome of *P. peltata* (NC052862) as a reference sequence. The primers for junction validation are listed in Supplementary Table S1. The cp genome obtained in this study has been submitted to the NCBI database (www.ncbi.nlm.nih.gov) with the GenBank accession number listed in Table 1. The physical maps of the cp genomes were produced with Chloroplot (<https://irscope.shinyapps.io/Chloroplot/>).

Table 1
Summary of features of four *Pinellia* chloroplast genomes.

Taxon	Accession number	Length (bp)				Number of genes				GC content (%)			
		Genome	LSC	SSC	IR	Total	Protein coding	tRNA	rRNA	Genome	LSC	SSC	IR
<i>P. pedatisecta</i>	MZ702636	164,682	90,947	22,523	25,606	129	85	36	8	35.71	33.92	29.43	42.09
<i>P. pedatisecta</i>	MN046890	168,178	92,963	23,981	25,617	130	85	37	8	35.08	33.35	26.85	42.09
<i>P. peltata</i>	NC052862	164,923	90,089	24,871	24,981	130	86	36	8	36.51	34.53	31.77	42.44
<i>P. ternata</i>	KR270823	164,013	89,783	22,980	25,625	131	86	37	8	36.66	34.6	31.66	42.53

Repeat structure identification

SSRs were identified by MISA [28], and the minimum thresholds for mono-, di-, tri-, tetra-, penta- and hexa-nucleotides were set to 10, 6, 5, 5, 5 and 5, respectively. REPuter was used to detect two kinds of long repeats: forward and palindromic repeats [29]. Detection parameter settings were used as follows: repfind -d -p -h 3 -l 30 -best 50. Tandem Repeats Finder (<http://tandem.bu.edu/trf/trf.html>) was used to find tandem repeats with the default settings.

Genome comparative analysis

In addition to the newly sequenced cp genome, 26 available cp genome sequences of *Pinellia* and related species were downloaded from the NCBI database: *P. pedatisecta* (MN046890), *P. ternata* (KR270823), *P. peltata* (NC052862), *Arisaema franchetianum* (MN046885), *Arisaema ringens* (MK111107), *Arisaema erubescens* (MT676834), *Arisaema nepenthoides* (MW338731), *Typhonium blumei* (NC051872), *Sauromatum giganteum* (NC050648), *Alocasia navicularis* (MN046882), *Colocasia esculenta* (JN105689), *Pistia stratiotes* (MN885890), *Amorphophallus konjac* (MK611803), *Calla palustris* (MN046887), *Epipremnum aureum* (KR872391), *Epipremnum amplissimum* (MN477424), *Aglaonema costatum* (MN046881), *Monstera adansonii* (MN046888), *Zantedeschia aethiopica* (KY792991), *Pothos scandens* (MN046891), *Symplocarpus renifolius* (KY039276), *Symplocarpus nipponicus* (MK341566), *Lemna minor* (DQ400350), *Acorus calamus* (AJ879453), *Acorus americanus* (EU273602) and *Acorus gramineus* (KP099646). The multiple sequence alignment of the 27 cp genome sequences was performed using MAFFT v.7 with the default settings and adjusted manually where necessary with BioEdit v.7.2.5 software. On the basis of the aligned sequence matrix of the cp genomes of *Pinellia*, interspecific nucleotide diversity (K) was evaluated by sliding window analysis with a step size of 1000 bp and window length of 2000 bp in DnaSP v.5.10 [30]. The evolutionary divergences of the four *Pinellia* cp genomes were evaluated using nucleotide differences and p-distance by MEGA v.10.0.4. The protein-coding sequences (CDSs) were extracted using Geneious v.2019.1.3 [31].

Phylogenetic analyses

Two datasets (cp genome and coding sequence) were used to construct the phylogenetic topology of 27 *Pinellia* and related species with maximum likelihood (ML) methods, respectively. ML analyses were performed using RAxML 8.2.9 under the GTRGAMMA model with 1000 rapid bootstrap replicates. Three species of *Acorus* (*A. calamus*, *A. americanus* and *A. gramineus*) were used as outgroups.

Divergence time estimation

Accurate estimation of the divergence time in a taxon is important to understand its evolutionary history. Divergence times were estimated using PAML mcmctree (PAML v.4.9j) [32] with the approximate likelihood calculation method. The analysis was performed on 27 complete cp genome sequences used in the phylogenetic analysis with three known calibration times: (1) divergence between *P. pedatisecta* and *T. blumei* was 36–32 million years ago (Mya); (2) divergence between *A. navicularis* and *P. stratiotes* was 85–46 Mya; and (3) all species except the genus *Acorus* from the Araceae family arose 122–117 Mya inferred from the published knowledge-based TimeTree (timetree.org). Posterior distributions of parameters were approximated using two independent mcmctree analyses of 10,000,000 generations with 20% burn-in. Tracer v.1.4.1 was used for checking the convergence of the chains through adequate effective sample sizes (ESSs).

Gene selection site analysis

A total of 74 single-copy protein-coding genes shared by four *Pinellia* plastid genomes (Table 1) were extracted and aligned by Geneious v.9.0.5 [31] and MAFFT v.7. The ML tree was constructed using RAxML V.8.2.9 with the GTRGAMMA model based on complete cp genomes. Protein-coding exon and each value of dN, dS and ω were calculated using the site-specific model in the Codeml program (seqtype = 1, model = 0, Nsites = 0,1,2,3,7,8) of PAML v.4.9j [32]. To determine the selected sites, we compared the model M0 (one ratio) versus M3 (discrete), M1 (neutral) versus M2 (positive selection) and M7 (beta) versus M8 (beta and ω) and carried out the three likelihood ratio tests (LRTs). Only consistent sites of positive selection with significant support from posterior probability (p of $(\omega \geq 1) \geq 0.99$; Bayes Empirical Bayes approach [BEB]) were identified. BEB recognized by Models M2 and M8 were further considered.

Results

Chloroplast genome features in *Pinellia*

In our trial, another cp genome of *P. pedatisecta* was sequenced, assembled and validated (Fig. S2). A total of 49,904,104 PE raw reads were generated using the Illumina Sequencing System. The novel cp genome sequence has been preserved in GenBank (MZ702636, Table 1). The cp genome of *P. pedatisecta* was circular double-stranded DNA and displayed a quadripartite structure (Fig. 1). The cp genome assembled in this study was 164,682 bp in length, which was 3,496 bp shorter than the published one (MN046890, 168,178 bp) mainly due to the contraction of the LSC region. Moreover, the GC content in the SSC region of the cp genome MZ702636 was 2.58% higher than that of MN046890. The cp genome MN046890 had the same rRNA gene content as that of MZ702636 but contained one more tRNA gene (*tRNA-His*) (Supplementary Table S2).

After downloading from the NCBI database, protein-coding gene number variations of 24 cp genomes of the Araceae family, including four *Pinellia* cp genomes, were analysed (Fig. 2). Compared with the cp genome of the model plant *Arabidopsis thaliana*, two genes, *infA* and *ycf68*, were inserted among most Araceae plants, while one-copy *ycf1* genes were missing. In the Araceae family, there were significant differences in the number of *ycf68* genes among species. The *ycf68* genes in 60% of Araceae cp genomes were completely missing, while the remaining 40% of the cp genomes had two copies. Even within the *Pinellia* genus, the number of *ycf68* genes was also inconsistent. Nevertheless, there was no discrepancy in the type and number of protein-coding genes between our newly assembled cp genome (MZ702636) and the published one (MN046890).

SSR and repeat sequence analyses

The number of SSRs in the four *Pinellia* cp genomes ranged from 68 (*P. peltata*) to 111 (*P. ternata*, Fig. 3A and Supplementary Table S3). Three kinds of SSRs were discovered, namely, mononucleotide, dinucleotide and trinucleotide. Among each *Pinellia* species, mononucleotide repeats were the most common, whereas trinucleotide repeats accounted for the lowest proportion of SSRs. The number of A/T mononucleotide repeats exceeded that of the other three types combined (Fig. 3A). Interspersed repeated sequences were identified by using REPuter for four plastomes (Fig. 3B and Supplementary Table S4). Except for IR regions, the repeats' length ranged from 31 bp to 328 bp with forward (F) repeats as the most prevalent type. Most interspersed repeats of two *P. pedatisecta* cp genomes were located in the LSC regions, while the long repeats of *P. peltata* and *P. ternata* were predominantly in the IRb/SSC and SSC regions, respectively. The total number of tandem repeats for four plastid genomes was in the range of 34–274. *P. peltata* had the least repeats, and *P. pedatisecta* (MN046890) had the most (Supplementary Table S5).

Comparative genomic analyses

The differences and evolutionary divergences among four *Pinellia* cp genomes were compared using nucleotide substitutions and sequence distance (Table 2). Across all four *Pinellia* cp genomes, the p-distance was 0.000402–0.013668, and the value of nucleotide differences was 66–2459. The comparison results of nucleotide substitution and genetic distance were consistent. The sequence divergence level of two cp genomes of *P. pedatisecta* was the lowest. The differentiation between *P. pedatisecta* and *P. ternata* was less than that between *P. peltata*. Among three *Pinellia* species, *P. ternata* and *P. peltata* were the most genetically distant. Moreover, the K value (sequence divergence between species) was calculated, and the sliding windows of the K values were constructed by DnaSP (Fig. 3C and Supplementary Table S6). Fig. 3C showed that the sequence divergence between *P. pedatisecta* and *P. ternata* was much lower than the two other K values. The divergences between *P. pedatisecta* and *P. peltata* and between *P. ternata* and *P. peltata* were not significant. Eight highly variable regions with great K values were detected, namely, *psbl-trnG*-

UCC, *psbM-rpoB*, *ndhJ-trnT-UGU*, *trnP-UGG-trnW-CCA*, *ndhF-trnN-GUU*, *ndhG-ndhE*, *ycf1-rps15* and *trnR-ycf1*. Four of these regions were located in the LSC region, and the remaining four were located in the SSC region, all of which were present in the non-coding regions.

Table 2

Numbers of nucleotide substitutions and sequence genetic distance in four complete cp genomes. The upper triangle shows the number of nucleotide substitutions and the lower triangle indicates the sequence distance.

	<i>P. pedatisecta</i> MZ702636	<i>P. pedatisecta</i> MN046890	<i>P. peltata</i> NC052862	<i>P. ternata</i> KR270823
<i>P. pedatisecta</i> MZ702636		0.000402	0.009860	0.007206
<i>P. pedatisecta</i> MN046890	66		0.011608	0.008599
<i>P. peltata</i> NC052862	1479	1778		0.013668
<i>P. ternata</i> KR270823	1275	1348	2459	

Phylogenetic analyses

To analyze the phylogenetic relationship of *Pinellia* species, we constructed phylogenetic trees using whole cp genome sequences and their CDSs by ML methods. For *Pinellia* species, the topological structure obtained by either complete cp genome or CDSs was roughly identical (Fig. 4). Both phylogenetic trees showed that *Pinellia* plants were monophyletic and clustered into two clades with strong support ([both whole cp genome and CDS datasets] bootstrap = 100). One clade was robustly supported ([both two datasets] bootstrap = 100) and comprised *P. pedatisecta* and *P. ternata*. The other clade consisted of one single species *P. peltata*. Among these three *Pinellia* species, *P. pedatisecta* and *P. ternata* were related. Compared with other plants in Araceae, the *Arisaema* genus was a sister group of *Pinellia* in our phylogeny based on either whole cp genome or CDSs (Fig. 4).

Divergence time estimation

In this study, we used full-length sequences of cp genomes of 23 Araceae family plants (including four *Pinellia* cp genomes) and three outgroups to estimate the divergence times of major clades in the Araceae family. Our dating analysis resulted in estimates for the crown node of the Araceae family of 91.99 Mya (95% highest posterior density [HPD] = 46.69–125.17 Mya; node 1 in Fig. 5; Supplementary Table S7) in the early Cretaceous. The divergence time of the four genera *Pinellia*, *Arisaema*, *Sauromatum* and *Typhonium* was estimated at 25.65 Mya (95% HPD = 12.93–36.35 Mya; node 2) in the Oligocene. The age estimation for the crown node of *Pinellia* and most species of *Arisaema* was dated to 22.48 Mya (95% HPD = 11.39–33.32 Mya; node 3) in the late Oligocene and early Miocene. *Pinellia* diverged from its sister clade at 13.09 Mya (95% HPD = 4.91–21.96 Mya; node 4 in Fig. 5; Supplementary Table S7) in the Miocene.

Gene selection pressure analysis of protein sequences

The site-specific selective pressure on four *Pinellia* plastomes was assumed using the site model in PAML program. Three pairs of site model comparisons (M0 vs M3, M1 vs M2a and M7 vs M8) showed four protein-coding genes subjected to positive selection (LRT of the three comparisons all $p < 0.05$; Supplementary Tables S8 and S9), namely, *petB*, *rpoC1*, *rps12* and *ycf1*. Of these positive selection genes, *ycf1* had the highest number of sites (21 sites; Table S8), followed by *rpoC1* (4 sites), *petB* (1 site) and *rps12* (1 site).

Discussion

Phylogenetic position of the *Pinellia* genus in Araceae

In the previous classification systems, the *Pinellia* genus belonged to the Araceae family, but its sister group remains controversial. In the Schott system [9], which is the first classification system of the Araceae family, the genus *Pinellia* was placed in the tribus Alleluhieae, related to two genera of *Cryptocoryne* and *Langenandra*. In the Engler system [10], *Pinellia* belonged to the subtrib. Pinelliinae, trib. Areae, subfam. Aroideae, in the same tribus as the *Arisaema* genus. In the Hutchinson system [11], this genus was placed in trib. Areae, related to two genera of *Cryptocoryne* and *Ambrosina*. In the classification system of Grayum [12], *Pinellia* belonged to trib. Pinellia, subfam. Aroideae. After modifying the Engler system by Bogner and Nicolson [13], the *Pinellia* genus changed to the subtrib. Atherurinae, trib. Pinellia, subfam. Aroideae. Mayo et al. [14] conducted taxonomic analysis of 106 taxa in the Araceae family using 63 morphological and anatomical data and proposed the latest classification system of the Araceae family. In this system, *Pinellia* was placed in trib. Arisaemateae, subfam. Aroideae with the *Arisaema* genus as sister group. Since the 21st century, with the development of molecular systematics, many studies have investigated the phylogenetic position of the *Pinellia* genus in Araceae [7, 15–18], yet controversy regarding molecular systematics and morphological classification persists. French et al. [16], Cabrera et al. [7] and Cusimano et al. [15] performed molecular phylogenetic analyses by applying different types and numbers of DNA marker sequence data, and they discovered that the *Arisaema* genus is strongly related to *Pinellia* among Araceae plants and is its sister group. Contrary to the above classification, Keating [17] and Bogner and Petersen [18] collected morpho-anatomical data; they argued that *Arisaema* and *Pinellia* cannot be

grouped into a unique clade. One significant morphological discrepancy between them was that almost all *Arisaema* species have at least partially fused stamens, whereas *Pinellia* and other related genera (e.g. *Sauromatum* and *Typhonium*) have free stamens.

The cp genome is one of three subcellular compartments in the plant genome, and it is mainly inherited from the maternal parent [33]. Given its high conservation and abundant interspecific variation, the cp genome has the potential to provide distinguishing differences that can help molecularly classify closely related species [34]. With advances in high-throughput sequencing, achieving the cp genome is easily acquirable at a large scale with low costs. Researchers have proposed the entire cp genome as a super barcode to discriminate and classify closely related species [35]. To date, phylogenies of several genus-level taxa have been further clarified by using cp genome sequences, such as *Epimedium* [36], *Paris* [37] and *Sanguisorba* [38]. Hence, we compared 24 cp genomes of the Araceae family, including four plastomes of *Pinellia*, to explore the cp genome molecular phylogeny of *Pinellia*. The resulting phylogeny based on either entire cp genomes or CDSs in this study showed that *Pinellia* and *Arisaema* plants were gathered into one branch and sister groups to each other with a well-supported bootstrap value (100%; Fig. 4). Although the Araceae family is a major group of monocotyledons, there is still a limited number of cp genomes available from Araceae species, which may result in some congeneric species, such as *Arisaema* plants, not clustering together in one phylogenetic branch (Fig. 4). The absence of plastid genomes from three other candidates of the *Pinellia* sister groups, *Cryptocoryne*, *Langenandra* and *Ambrosina* [21], could also hinder our discovery of the intergenus phylogeny of *Pinellia*.

Phylogeny within the *Pinellia* genus

Despite recent advances in molecular phylogenetic studies, deep evolutionary relationships and below-genus taxonomic classification of *Pinellia* remain unresolved. Yin [21] made the first comprehensive phylogenetic analyses of all seven *Pinellia* plants by using the sequences of ITS and *trnL-F* DNA barcodes. The resulting phylogeny of both ITS and *trnL-F* sequences supported that *P. pedatisecta* and the six remaining *Pinellia* species were sister groups and served as a basal taxon in *Pinellia*. Moon et al. [39] reconstructed the phylogenetic trees of three species, *P. ternata*, *P. tripartita* and *P. pedatisecta*, with *matK* and *rbcL* sequences; they found that the topology of two trees is inconsistent. The ancestor showed in the *matK* tree was *P. pedatisecta*, congruent with the phylogeny above by Yin [21], while the ancestor in the *rbcL* tree was *P. ternata*. Furthermore, after morphological summarization, Yin found that *P. pedatisecta* has many characteristics of the *Pinellia* genus. These traits include perennial herb with a small tuber at the top of the main tuber, persistent spathe, female inflorescence adnate to the spathe, unisexual flowers, no perianth, one straight ovule and green fruit. Hence, combined with the molecular phylogeny, Yin concluded that the basal taxon of the *Pinellia* genus is *P. pedatisecta*, one species native to shady woodland areas, forested slopes and valleys in northern and western China; he suggested that the taxonomic rank of *P. pedatisecta* should be elevated to a section of *Pinellia* [21], which was the first and only below-genus taxonomic recommendation of *Pinellia*.

In our opinion, *P. pedatisecta* indeed has some of the characteristic traits of *Pinellia*, but it also possesses so many characteristics significantly different from other congeneric species. For example, lanceolate spathe, no constriction in the tube and eaves of spathe, no diaphragm between the female inflorescence and male inflorescence and no bulblet on the petiole [15, 19, 20]. Our group previously conducted phylogenetic analyses of *Pinellia* by using four different barcode sequences with maximum parsimony (MP) method. The topologies of the four barcodes were not consistent, with poor bootstrap values (Fig. S1) probably due to insufficient sequence length and interspecific variation. Hence, plastid genomes, an ideal model for evolutionary and comparative genomic studies of related species, were necessary here for a molecular phylogeny of *Pinellia* with a significantly higher resolution.

After comparing the structural organization of three published cp genomes of *Pinellia*, we found that the sequence of *P. pedatisecta* (MN046890) was not highly similar to those of two other species, *P. peltata* (NC052862) and *P. ternata* (KR270823). (1) The length of the *P. pedatisecta* plastome was significantly longer than that of the two other species (Table 1), specifically 3.2 kb longer than that of *P. peltata* and even 4.1 kb longer than that of *P. ternata*. The length of the LSC region in *P. pedatisecta* was 3.18 kb longer than that of *P. ternata*. (2) With regard to protein-coding gene content, all two *ycf68* genes were missing in the published cp genome of *P. pedatisecta* (Fig. 2), compared with *P. peltata* and *P. ternata*, which could have the potential to affect the physiological functions of plants, although some authors suggested that the *ycf68* gene likely does not encode a protein [40]. (3) The GC content of published *P. pedatisecta* was lower than that in *P. peltata* and *P. ternata* (Table 1), especially in the SSC region, which was 4.92% lower than that in *P. peltata* and 4.81% lower than that in *P. ternata*. (4) After the alignment of three published cp genomes and further estimation of locally collinear blocks (LCBs) with MAUVE 2.4.0, the gene order comparison revealed one rearrangement (~128 bp) between the plastomes of *P. pedatisecta* and *P. ternata* (Fig. S3A and S3B). (5) A significant insertion/deletion variation was noted across three cp genomes in this group, located between *tRNA-Ser (GCU)* and *tRNA-Ser (CGA)* genes in the LSC region with the length of ~1.2 kb (10714–11950 bp in the published *P. pedatisecta* cp genome and 10192–10234 bp in the self-assembled one; Fig. S3C and S3D), all five of which do not commonly occur, or are relatively rare, in conserved plastid genomes of congeneric species. Therefore, we determined to identify, sample and sequence the plastome of *P. pedatisecta* for further phylogenetic analyses of *Pinellia*.

In this study, phylogenetic trees based on both datasets (whole cp genomes and CDSs) showed that the species of *Pinellia* clustered into two clades with high support values (Fig. 4). One clade comprised *P. pedatisecta* and *P. ternata*, which possess pedate and three full-lobed leaves, respectively, with strong support. The other clade was robustly supported and comprised one species, *P. peltata*, characterized by undivided and peltate leaf and development of stem tuber. *P. pedatisecta* is buried inside the cp genome phylogeny, incongruent with the phylogeny from ITS and *trnL-F* sequences

by Yin [21]. Furthermore, comparative genomic analyses showed that the nucleotide substitutions and sequence distance between *P. pedatisecta* and *P. ternata* were the smallest (1275 and 0.007206; Table 2), while that between *P. ternata* and *P. peltata* was the largest (2459 and 0.013668). The interspecific nucleotide diversity (K value) between *P. pedatisecta* and *P. ternata* was considerably lower than the two other K values (Fig. 3C), both of which confirmed that *P. peltata* was relatively genetically distant from *P. pedatisecta* and *P. ternata*. Limited by the number of *Pinellia* cp genomes published so far, we cannot determine whether *P. peltata* is the ancestor of *Pinellia* plants, but it is possible. *P. peltata* has the same floral morphological characteristics as other *Pinellia* species, except for *P. pedatisecta* [25]. For example, solitary inflorescence, persistent spathe, with constriction in the tube and eaves of spathe, female inflorescence adnate to the spathe, with diaphragm between the female inflorescence and male inflorescence, unisexual flowers, no perianth, one locular ovary and one straight ovule [21]. However, the peltata leaf characteristics of *P. peltata* are considerably different. Thus, the below-genus taxonomic classification of *Pinellia* is suggested to be based on differences in leaves or other tissue morphological characteristics between species, rather than flowers.

Variations and evolution of *Pinellia* cp genomes

In this study, the cp genome of *P. pedatisecta* was sequenced (MZ702636) with the length of 164,682 bp, which fell within the cp genome size range for angiosperms but much smaller than its published one (MN046890, 168,178 bp; Table 1). Among four *Pinellia* plastid genomes, LSC regions showed the most difference in size, with the shortest 89,783 bp of *P. ternata* KR270823 and the longest 92,963 bp of *P. pedatisecta* MN046890. Additionally, the inferred structures and protein-coding gene contents were in accordance except for the *infA* and *ycf68* genes (Fig. 1 and 2). Analysis with DnaSP inferred that some of the most divergent regions of *psbI-trnG-UCC*, *psbM-rpoB*, *ndhJ-trnT-UGU*, *trnP-UGG-trnW-CCA*, *ndhF-trnN-GUU*, *ndhG-ndhE*, *ycf1-rps15* and *trnR-ycf1*, as shown in Fig. 3C, were found for further related species identification of *Pinellia*.

SSRs in the cp genome are an efficient marker tool for population genetic structure and phylogeography [41]. In this study, 68–111 SSR loci were identified between *Pinellia* species (Supplementary Table S3). These SSR loci could provide candidate molecular markers for the genetic diversity study of *Pinellia*. The composition of SSR loci in the cp genomes of three *Pinellia* species was similar to that of most angiosperms, with A/T mononucleotide repeats dominating all the repeat units. This phenomenon may be one of the reasons for the abundance of A/T bases in cp genomes. Repeat units in the cp genome can cause sequence polymorphism, and long repeats in the cp genomes of *Pinellia* are abundant in the forward and palindromic repeats (Supplementary Table S4).

At present, specific studies on the divergence time estimation of *Pinellia* by using molecular data, particularly plastid genomes, are lacking. Li et al. [42] first supposed that species with simple leaves (e.g. *P. peltata*) are less evolved than those with pedate or lobed leaves (e.g. *P. pedatisecta* and *P. ternata*) on the basis of the palynology and isozyme characteristics of *Pinellia* plants. Yin [21] speculated that the *Pinellia* genus first appeared in the Paleogene according to geographic distribution and the origin time of its related genus (*Arisaema*). Here, based on whole cp genome sequences, our divergence time estimation within Araceae indicated that *Pinellia* species originated at ~22.48 Mya (95% HPD = 11.39–33.32 Mya; node 3 in Fig. 5; Supplementary Table S6) in the late Oligocene and early Miocene, congruent with that estimated by Yin and further refined, and diverged diversely at ~13.09 Mya (95% HPD = 4.91–21.96 Mya; node 4 in Fig. 5) in the Miocene. Furthermore, within the *Pinellia* genus, *P. peltata* diverged first at ~5.34 Mya earlier than two other species, *P. pedatisecta* and *P. ternata*; these findings were consistent with the results of Li et al [42]. The present estimation of divergence times based on plastid genomes and fossil data provides new insights and a hypothetical foundation for future studies on the origin and earlier evolution of *Pinellia*.

Positive selection genes have a significant impact on the creative effects of populations, changes under selection stress, and genetic drift led to the rapid transformation of genes into new common adaptive combinations [43]. In this study, we calculated the non-synonymous/synonymous substitution rate ratio ($\omega = dN/dS$) for each of the 74 single copy protein-coding genes shared by the analyzed plastid genomes of *Pinellia*. Four genes with high posterior probability of codon sites in the BEB test were acquired and considered as positive selection genes (Fig. 6 and Tables S8 and S9). These genes included one gene for cytochrome b/f complex subunit protein in the photosystem II reaction (*petB*) [44], one DNA-dependent RNA polymerase gene (*rpoC1*), one gene for ribosome small subunit protein (*rps12*) and *ycf1* gene. All these positive selection genes were also detected in other plants [45–48], and these genes may have played a significant role in the adaptive evolution of *Pinellia*. The specific role needs to be further studied.

The genus *Pinellia* includes seven species in total. Only three *Pinellia* plants were studied here, and as a consequence, it might not fully show the real effect of the phylogeny of *Pinellia*. However, our study was the first comprehensive phylogenetic analysis of *Pinellia* at the cp genome level, and the resulting topologies based on either whole cp genome or CDS datasets both definitively show that *P. pedatisecta* cannot be the ancestor of *Pinellia* plants. The main habitat of all three *Pinellia* species in our study is in the southeastern portion of Asia, where hybridization and introgression have been reported [49, 50], which may cause interspecific gene flow in response to ecological selection and lead to difficulties in molecular phylogenetic analysis. Given that the cp and nuclear genomes evolve independently, the phylogeny of cp genomes alone is insufficient in making taxonomic decisions about *Pinellia*. Therefore, more methods and more species are needed in further study to improve our ability to better understand the phylogeny and evolution of *Pinellia*.

Conclusions

The genus *Pinellia* Tenore is comprised of seven species, all of which are important medicinal plant resources. Species phylogenetic classification is vital for protecting species diversity and selecting high quality germplasm resources. But due to members of *Pinellia* exhibit rather wide variations in morphological structures, the low-level taxonomy and interspecific phylogeny of *Pinellia* are difficult to address based on morphology. With the development of next-generation sequencing technology, complete chloroplast genomes have been widely employed to explore phylogenetic relationships of intra- or inter-genus. However, the variation and evolution of the whole chloroplast genomes in the genus *Pinellia* have been ignored. Here, we sequenced the chloroplast genome of *P. pedatisecta*, compared it with previously published plastid genomes, and reconstructed phylogenetic relationships of *Pinellia* based on chloroplast genomes and their CDSs. The divergence time estimation and selective pressure of *Pinellia* plastomes were also investigated. The results showed some variations and adaptive evolution between *Pinellia* complete chloroplast genomes including size, structure and nucleotide diversity, which provided valuable information for species classification and evolution. In addition, our results revealed a sister relationship between *Pinellia* and *Arisaema* plants whose divergence was estimated to occur around 22.48 million years ago. All *Pinellia* plants formed a monophyletic evolutionary clade in which *P. peltata*, rather than *P. pedatisecta*, earlier diverged, indicating that *P. pedatisecta* is not the basal taxon of *Pinellia* but *P. peltata* may be. In conclusion, our results could provide insight into the chloroplast genome evolution and phylogeny of genus *Pinellia* and even of family Araceae, which would be useful for selecting high quality *Pinellia* germplasm resources in the future.

Declarations

Acknowledgements

The authors thank Baosheng Liao for his support in data analysis.

Author contributions

NC and XL conceived and designed the study; NC and WC collected and analyzed the data; NC drafted the initial version of the manuscript; NC, XL and PW contributed to later versions of the manuscript.

Funding

This study was supported by the National Key Research and Development Program of China (No.2019YFC1710600), Research Incubation Fund Project of Shandong Academy of Chinese Medicine (2021SACM-3), Major Science and Technology Innovation Projects of Shandong Province (2018CXGC1310), and the Natural Foundation of Shandong Province (ZR2019MH134).

Availability of data and material (data transparency)

The chloroplast genome sequence of *Pinellia pedatisecta* assembled here is accessible via GenBank with the accession number of MZ702636 and Global Pharmacopoeia Genome Database (GPGD,

<http://www.gpggenome.com/species/40314>). Raw sequencing data is available at NCBI SRA database with the accession number of SRR15328795.

Compliance with ethical standards

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

References

1. Liu Y, Hui RK, Deng RN, Wang JJ, Wang M, Li ZY (2012) Abnormal male meiosis explains pollen sterility in the polyploid medicinal plant *Pinellia ternata* (Araceae). *Genet Mol Res* 11(1):112–120. <https://doi.org/10.4238/2012.January.17.1>

2. Zhu Y, Zhu G, Guo Q, Zhu Z, Wang C, Liu Z (2013) A comparative proteomic analysis of *Pinellia ternata* leaves exposed to heat stress. *Int J Mol Sci* 14(10):20614–20634. <https://doi.org/10.3390/ijms141020614>
3. Kozlov VA, Zhuravkin IN, Tsyrova IG (1978) [Comparative analysis of the migratory capacity of hematopoietic stem cells and of immunocompetent precursors after acute hypoxia and the administration of testosterone]. *Dokl Akad Nauk SSSR* 239(3):747–749
4. Nauheimer L, Metzler D, Renner SS (2012) Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytol* 195(4):938–950. <https://doi.org/10.1111/j.1469-8137.2012.04220.x>
5. Renner SS, Zhang LB (2004) Biogeography of the *Pistia* clade (Araceae): based on chloroplast and mitochondrial DNA sequences and Bayesian divergence time inference. *Syst Biol* 53(3):422–432. <https://doi.org/10.1080/10635150490445904>
6. Renner SS, Zhang LB, Murata J (2004) A chloroplast phylogeny of *Arisaema* (Araceae) illustrates Tertiary floristic links between Asia, North America, and East Africa. *Am J Bot* 91(6):881–888. <https://doi.org/10.3732/ajb.91.6.881>
7. Cabrera LI, Salazar GA, Chase MW, Mayo SJ, Bogner J, Davila P (2008) Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. *Am J Bot* 95(9):1153–1165. <https://doi.org/10.3732/ajb.0800073>
8. Boyce PC (1996) The Genus *Hapaline* (Araceae: Aroideae: Caladieae). *Kew Bull* 51(1):63–82. <https://doi.org/10.2307/4118745>
9. Boyce P, Schott H (1860) Schott, H.W. *Prodromus Systematis Aroidearum* (1860)
10. Engler A (1920) Araceae. Pars generalis et Index familiae generalis. In: Edited by Engler A, vol. 74 (IV. 23A). Leipzig, Berlin, Germany: Wilhelm Engelmann
11. Hutchinson J (1959) The families of flowering plants, arranged according to a new system based on their probable phylogeny, vol 2, 2nd edn. Internet Archive
12. Grayum MH (1990) Evolution and phylogeny of the Araceae. *Ann Mo Bot Gard* 77:628–697. <https://doi.org/10.2307/2399668>
13. Bogner J, Nicolson DH (1991) A revised classification of Araceae with dichotomous keys. *Willdenowia* 21:35–50
14. Mayo SJ, Bogner J, Boyce PC (1998) Araceae. In: *Flowering Plants Monocotyledons*. Edited by Kubitzki K. Berlin, Heidelberg: Springer Berlin Heidelberg:26-74
15. Cusimano N, Bogner J, Mayo SJ, Boyce PC, Wong SY, Hesse M, Hetterscheid WL, Keating RC, French JC (2011) Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies. *Am J Bot* 98(4):654–668. <https://doi.org/10.3732/ajb.1000158>
16. French JC, Chung MG, Hur YK (1995) Chloroplast DNA phylogeny of the Ariflorae. In: *Monocotyledons: Systematics and evolution*. Edited by Rudall PJ, Cribb P, Cutler DF, Humphries CJ, vol. 1: Royal Botanic Gardens, Kew, U.K.: 255-275
17. Keating RC (2004) Vegetative anatomical data and its relationship to a revised classification of the genera of Araceae. *Ann Missouri Bot Gard* 91:485–494
18. Fontana ME, Cavalcante CAV (2013) Índices baseados no número de clientes para localização de itens em armazéns. *Production* 23(3):561–569. <https://doi.org/10.1590/s0103-65132013005000002>
19. Li YW, Zhou FQ, Zhang SP, Li ZY, Kong QY (2008) [Studies on morphology and histology of *Pinellia pedatisecta*]. *Zhong Yao Cai* 31(2):206–209
20. Guo QS, Wang QY, Shi HZ (2006) [Study on morphological characteristics of pollen grains of *Pinellia ternata* in different populations]. *Zhongguo Zhong Yao Za Zhi* 31(1):27–30
21. Yin T: **Phylogeny and biogeography of *Pinellia* (Araceae) with special reference to phylogenetic problems of *Remusatia***. In., 2002 edn: Kunming Institute of Botany, Chinese Academy of Sciences; 2002
22. Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, McKain MR (2020) Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae). *Genomics* 112(3):2349–2360. <https://doi.org/10.1016/j.ygeno.2020.01.006>
23. Henriquez CL, Abdullah, Ahmed I, Carlsen MM, Zuluaga A, Croat TB, McKain MR (2020) Molecular evolution of chloroplast genomes in Monsteroideae (Araceae). *Planta* 251(3):72. <https://doi.org/10.1007/s00425-020-03365-7>
24. Zhai W, Duan X, Zhang R, Guo C, Li L, Xu G, Shan H, Kong H, Ren Y (2019) Chloroplast genomic data provide new and robust insights into the phylogeny and evolution of the Ranunculaceae. *Mol Phylogenet Evol* 135:12–21. <https://doi.org/10.1016/j.ympev.2019.02.024>
25. Luo R, Du S, Luo X (2021) The complete plastome sequence of *Pinellia peltata* Pei (Araceae). *Mitochondrial DNA B Resour* 6(8):2112–2113. <https://doi.org/10.1080/23802359.2021.1923411>
26. Cai Z, Wang H, Wang G (2020) Complete chloroplast genome sequence of *Pinellia ternata* (Thunb.) Breit, a medicinal plants to China. *Mitochondrial DNA B Resour* 5(3):2107–2108. <https://doi.org/10.1080/23802359.2020.1765207>
27. Cui N, Liao BS, Liang CL, Li SF, Zhang H, Xu J, Li XW, Chen SL (2020) Complete chloroplast genome of *Salvia plebeia*: organization, specific barcode and phylogenetic analysis. *Chin J Nat Medicines* 18(8):563–572. [https://doi.org/10.1016/S1875-5364\(20\)30068-6](https://doi.org/10.1016/S1875-5364(20)30068-6)
28. Beier S, Thiel T, Munch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33(16):2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>
29. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29(22):4633–4642. <https://doi.org/10.1093/nar/29.22.4633>

30. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
31. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
32. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591
33. McCauley DE, Sundby AK, Bailey MF, Welch ME (2007) Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris* (Caryophyllaceae): evidence from experimental crosses and natural populations. *Am J Bot* 94(8):1333–1337. <https://doi.org/10.3732/ajb.94.8.1333>
34. Wu L, Wu M, Cui N, Xiang L, Li Y, Li X, Chen S (2021) Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. *Chin Med* 16(1):52. <https://doi.org/10.1186/s13020-021-00460-z>
35. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. *Biol Rev Camb Philos Soc* 90(1):157–166. <https://doi.org/10.1111/brv.12104>
36. Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W, Zhang W, Kim K, Lee SC, Yang TJ et al (2016) The Complete Chloroplast Genome Sequences of Five *Epimedium* Species: Lights into Phylogenetic and Taxonomic Analyses. *Front Plant Sci* 7:306. <https://doi.org/10.3389/fpls.2016.00306>
37. Huang Y, Li X, Yang Z, Yang C, Yang J, Ji Y (2016) Analysis of Complete Chloroplast Genome Sequences Improves Phylogenetic Resolution in Paris (Melanthiaceae). *Front Plant Sci* 7:1797. <https://doi.org/10.3389/fpls.2016.01797>
38. Meng XX, Xian YF, Xiang L, Zhang D, Shi YH, Wu ML, Dong GQ, Ip SP, Lin ZX, Wu L et al (2018) Complete Chloroplast Genomes from *Sanguisorba*: Identity and Variation Among Four Species. *Molecules* 23(9):2137. <https://doi.org/10.3390/molecules23092137>
39. Moon BC, Kim WJ, Ji Y, Lee YM, Kang YM, Choi G (2016) Molecular identification of the traditional herbal medicines, *Arisaematis Rhizoma* and *Pinelliae Tuber*, and common adulterants via universal DNA barcode sequences. *Genet Mol Res* 15(1). <https://doi.org/10.4238/gmr.15017064>
40. Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK (2007) Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. *BMC Genom* 8:174. <https://doi.org/10.1186/1471-2164-8-174>
41. Angioi SA, Desiderio F, Rau D, Bitocchi E, Attene G, Papa R (2009) Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biol (Stuttg)* 11(4):598–612. <https://doi.org/10.1111/j.1438-8677.2008.00143.x>
42. Li L (1999) Micromorphological Characteristics and Enzyme Analysis of Some Chinese *Pinellias* and Their Taxonomic Significance. *Acta Botanica Yunnanica* 21(4):442–448
43. Xie DF, Yu Y, Deng YQ, Li J, Liu HY, Zhou SD, He XJ (2018) Comparative Analysis of the Chloroplast Genomes of the Chinese Endemic Genus *Urophysa* and Their Contribution to Chloroplast Phylogeny and Adaptive Evolution. *Int J Mol Sci* 19(7):1847. <https://doi.org/10.3390/ijms19071847>
44. Gui L, Jiang S, Xie D, Yu L, Huang Y, Zhang Z, Liu Y (2020) Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. *Gene* 732. <https://doi.org/10.1016/j.gene.2020.144355>
45. Liu ML, Fan WB, Wang N, Dong PB, Zhang TT, Yue M, Li ZH (2018) Evolutionary Analysis of Plastid Genomes of Seven *Lonicera* L. Species: Implications for Sequence Divergence and Phylogenetic Relationships. *Int J Mol Sci* 19(12):4039. <https://doi.org/10.3390/ijms19124039>
46. Chen Y, Zhong H, Zhu Y, Huang Y, Wu S, Liu Z, Lan S, Zhai J (2020) Plastome structure and adaptive evolution of *Calanthe* s.l. species *PeerJ* 8:e10051. <https://doi.org/10.7717/peerj.10051>
47. Gao C, Deng Y, Wang J (2018) The Complete Chloroplast Genomes of *Echinacanthus* Species (Acanthaceae): Phylogenetic Relationships, Adaptive Evolution, and Screening of Molecular Markers. *Front Plant Sci* 9:1989. <https://doi.org/10.3389/fpls.2018.01989>
48. Petersen G, Darby H, Lam VKY, Pedersen HAE, Merckx V, Zervas A, Seberg O, Graham SW (2019) Mycoheterotrophic *Epirixanthes* (Polygalaceae) has a typical angiosperm mitogenome but unorthodox plastid genomes. *Ann Bot-London* 124(5):791–807. <https://doi.org/10.1093/aob/mcz114>
49. Carlhoff S, Duli A, Nagele K, Nur M, Skov L, Sumantri I, Oktaviana AA, Hakim B, Burhan B, Syahdar FA et al (2021) Genome of a middle Holocene hunter-gatherer from Wallacea. *Nature* 596(7873):543–547. <https://doi.org/10.1038/s41586-021-03823-6>
50. Kinoshita G, Nunome M, Kryukov AP, Kartavtseva IV, Han SH, Yamada F, Suzuki H (2019) Contrasting phylogeographic histories between the continent and islands of East Asia: Massive mitochondrial introgression and long-term isolation of hares (*Lagomorpha*: *Lepus*). *Mol Phylogenet Evol* 136:65–75. <https://doi.org/10.1016/j.ympev.2019.04.003>

Figures

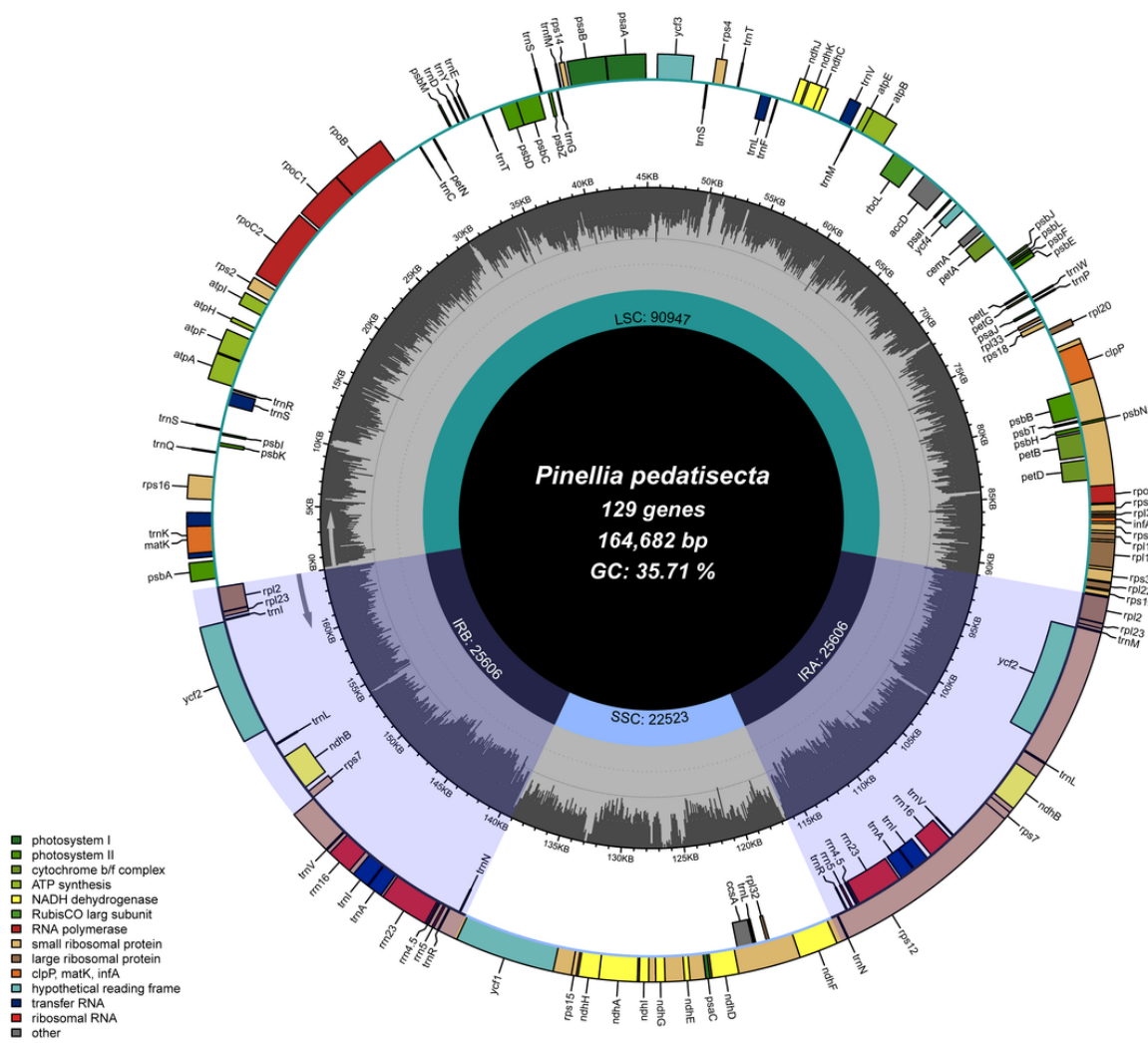


Figure 1
 Chloroplast genome map of *P. pedatisecta* assembled in this study. The centre of the figure provides length information of the cp genome. In the first inner circle, the proportion of the shaded parts represents the GC content of each part. The gene names are labelled on the outermost layer. The transcription directions for the inner and outer genes are listed clockwise and anticlockwise, respectively.

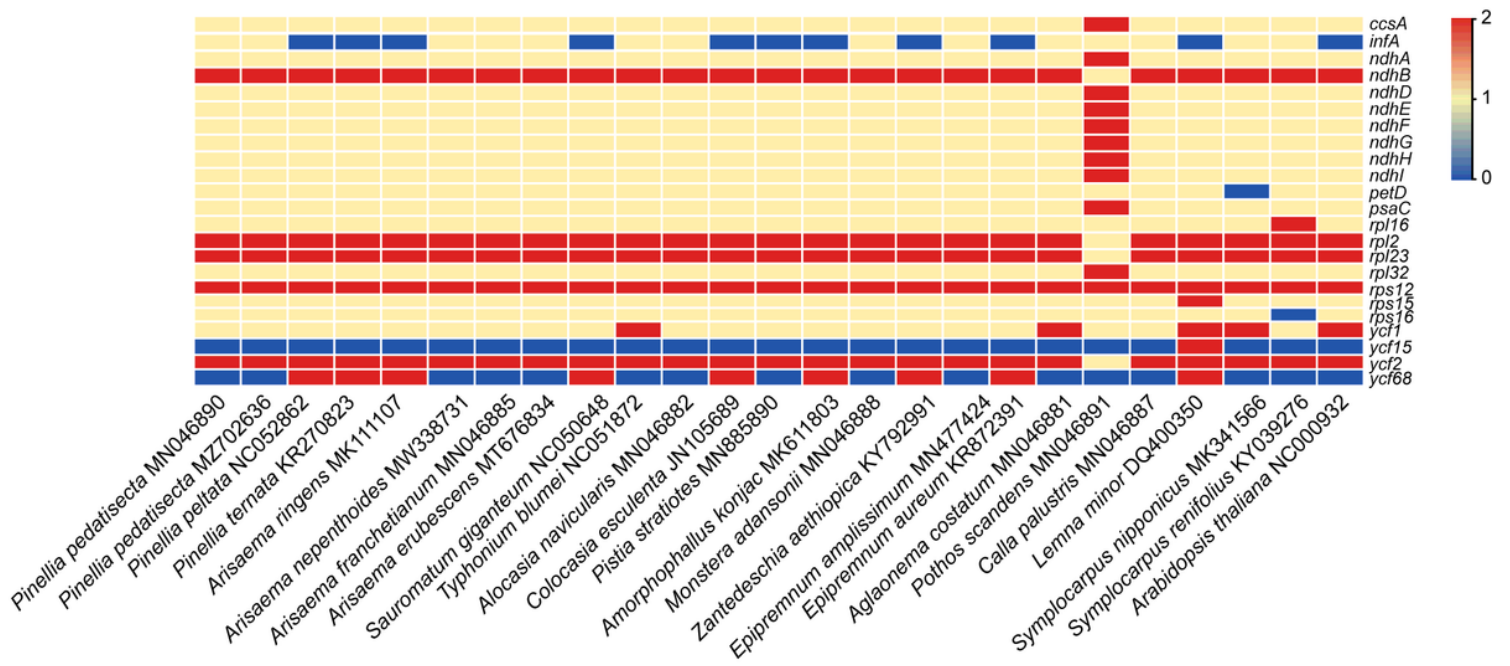


Figure 2

Comparison of the cp genome gene content of Araceae plants and one model plant (*Arabidopsis thaliana*). Only protein-coding genes with different numbers in 25 cp genomes are listed. Red boxes indicate two copies of each gene, and yellow boxes indicate a single copy of each gene. The blue boxes indicate an absent gene.

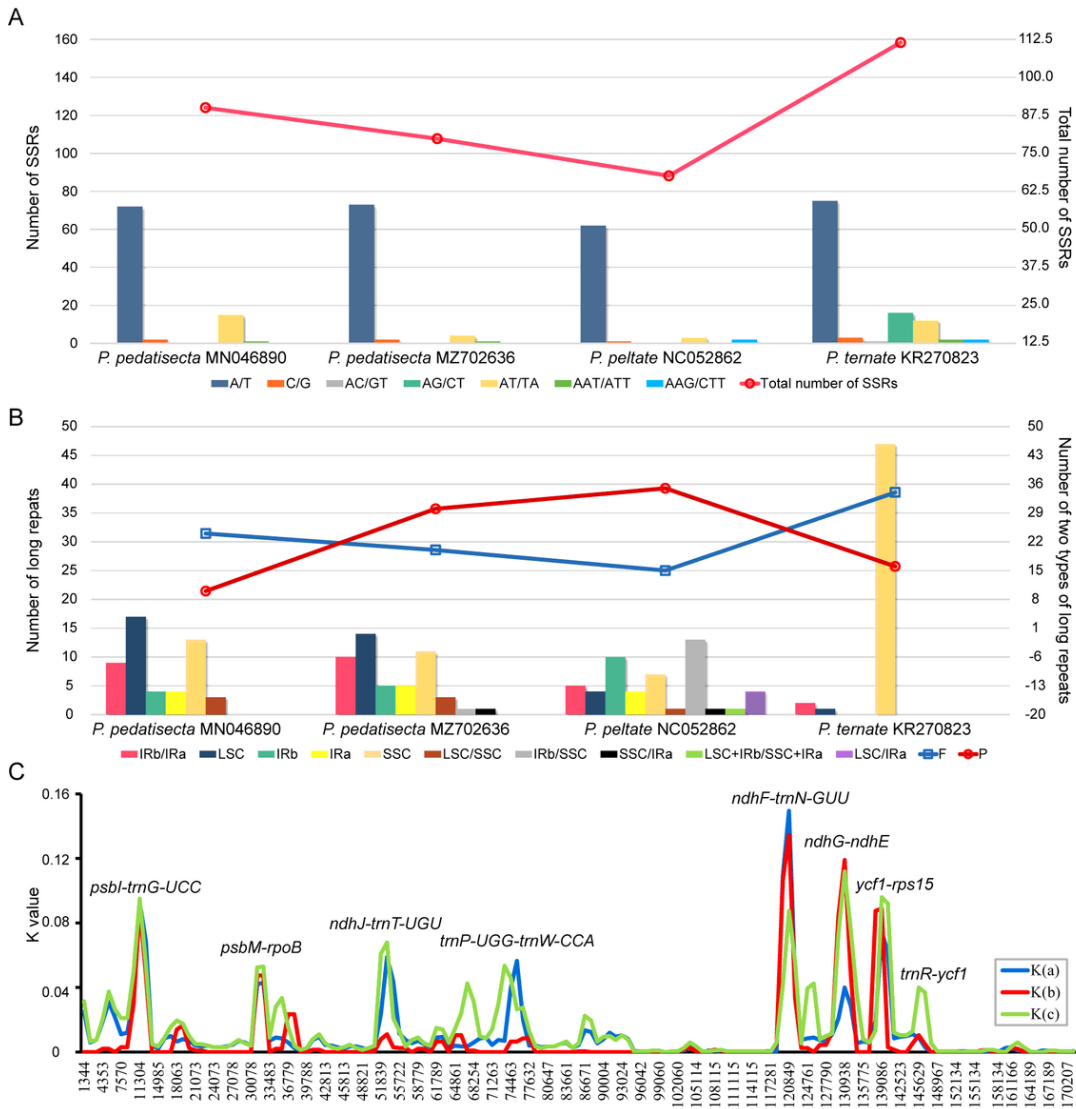


Figure 3

The SSRs (A), interspersed repeated sequences (B), and interspecific nucleotide diversity (C) among the chloroplast genomes of Pinellia. X-axis in Fig. 3C: position of a window. Y-axis in Fig. 3C: sequence divergence (K values) between species of each window. K(a): K values between *P. pedatisecta* and *P. peltata*; K(b): K values between *P. pedatisecta* and *P. ternata*; K(c): K values between *P. ternata* and *P. peltata*.

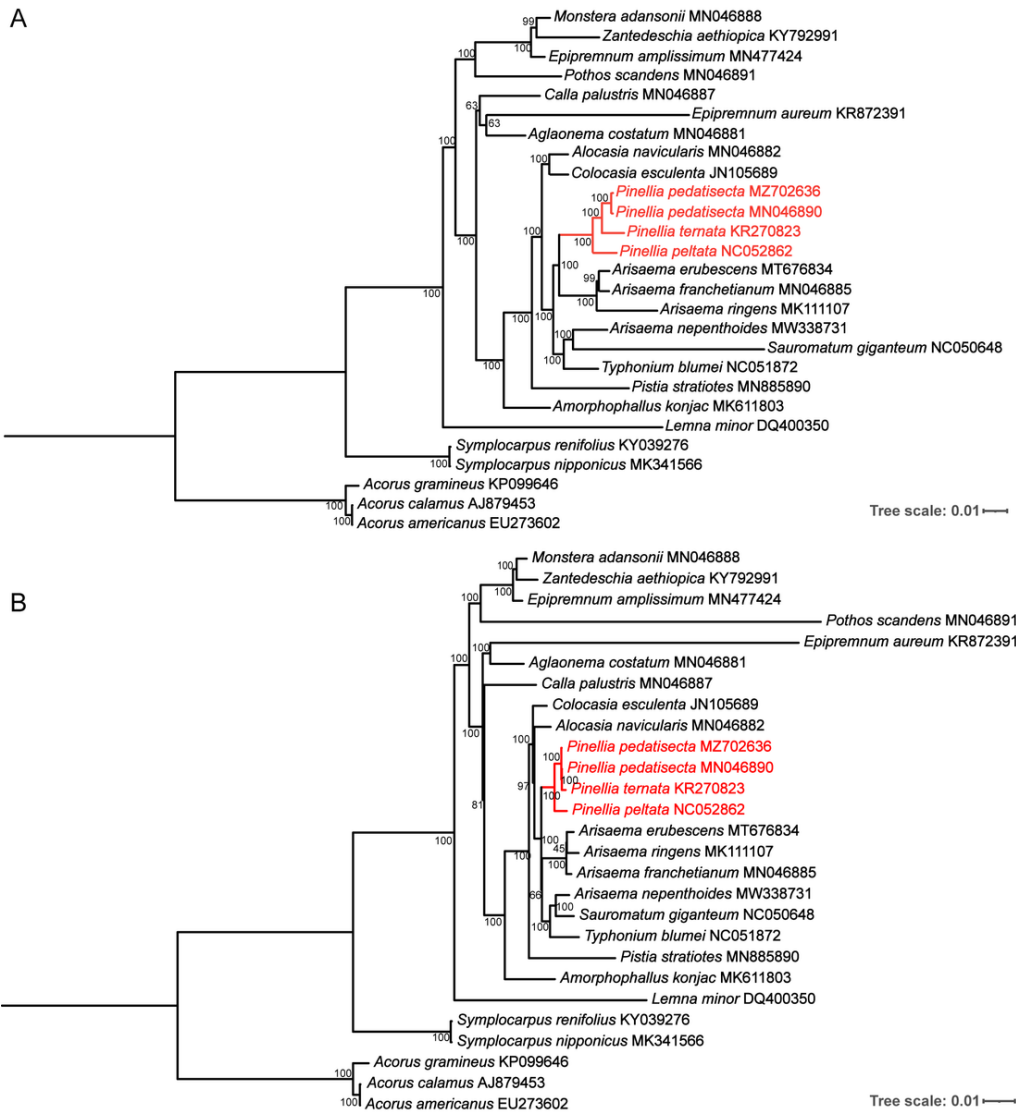


Figure 4

Phylogenetic trees constructed with the whole cp genomes (A) and their protein-coding sequences (B) by using ML method. Numbers near each branch are bootstrap values.

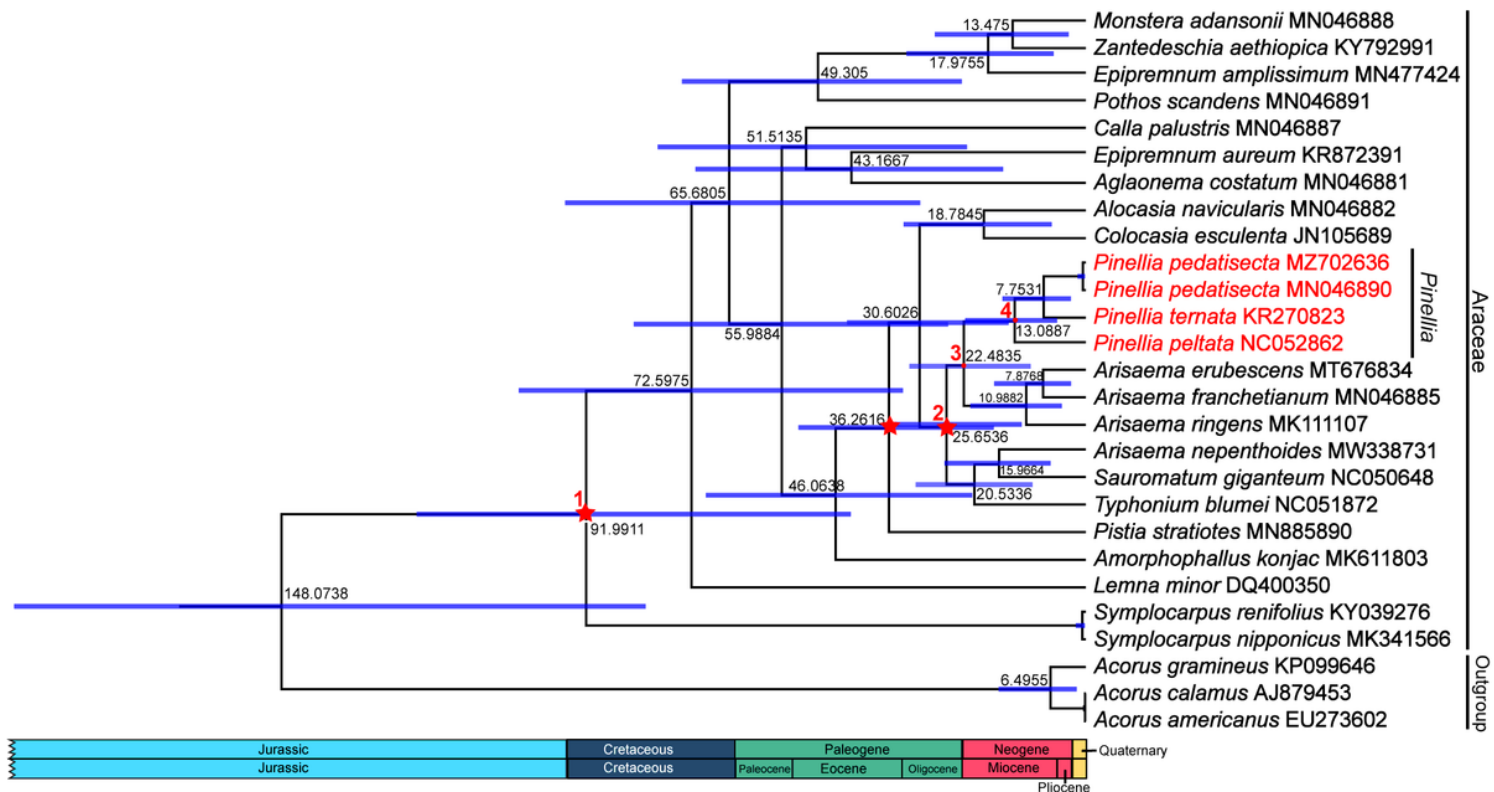


Figure 5
 Chronogram showing divergence times estimated based on 27 whole chloroplast cp genomes. Divergence time are shown for each node. Purple bars represented 95% HPD values for the estimated mean dates. Nodes labelled red pentagram were calibration points used for analysis. Numbers 1-4 indicated nodes of interest.



Figure 6
 Positive selection sites logo of four Pinellia plastomes

Supplementary Files

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

- [supplementFigures20210818.docx](#)
- [supplementTables0918.xlsx](#)