Comparative Chloroplast Genomics of Four *Pilea* Species (Urticaceae): High Levels of Sequences Divergence Provides New Insight into Interspecific Diversity in *Pilea*

Jingling Li  
Southwest University

Jianmin Tang  
Chongqing University of Arts and Sciences

Siyuan Zeng  
Southwest University

Fang Han  
Southwest University

Jing Yuan  
Southwest University

Jie Yu (✉ yujie1982@swu.edu.cn)  
Southwest University

Research article

**Keywords:** Pilea, chloroplast genome, interspecific diversity, identification, phylogenetic analysis

**Posted Date:** August 17th, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-49270/v1

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

**Version of Record:** A version of this preprint was published at BMC Plant Biology on January 7th, 2021. See the published version at https://doi.org/10.1186/s12870-020-02793-7.
Abstract

Background

Pilea is a genus of perennial herbs from the family Urticaceae, which are used for courtyard ornamental. For some species, they are used as medicinal plants in traditional Chinese medicine as well. The morphological characteristics of medicinal species from *Pilea* are similar, and it is difficult to accurately distinguish them based only on morphological characteristics. Besides, the species classification of *Pilea* are still controversial. The classification of many species are still in an unresolved state. At present, there is no information about the chloroplast genomes of *Pilea*, which limits our further understanding of this genus. Here, we first reported 4 chloroplast genomes of *Pilea* taxa (*P. mollis*, *P. glauca*, *P. peperomioides* and *P. serpyllacea*), and performed comprehensive comparative analysis.

Results

The four chloroplast genomes have similar structural characteristics and gene order with other angiosperms. These genomes all have a typical quartile structure, which contains 113 unique genes, including 79 protein-coding genes, 4 rRNA genes, and 30 tRNA genes. Besides, we detected SSRs and repeat sequences, and analyzed the expansion/contraction of IR regions. In particular, the comparative analysis showed a rather level of sequence divergence in the non-coding regions, even in the protein-coding regions of the four genome sequences, suggesting a high level of genetic diversity in *Pilea*. Moreover, we identified eight hypervariable regions, including *petN-psbM; psbZ-trnG-GCC; trnT-UGU-trnL-UAA; accD-psbl; ndhF-rpl32; rpl32-trnL-UAG; ndhA-intron* and *ycf1*, are proposed for use as DNA barcode regions. Phylogenetic analysis showed that four *Pilea* species form a monophyletic cluster with a 100% bootstrap value.

Conclusion

The results obtained here could provide abundant information for the phylogenetic position of *Pilea* and further species identification. High levels of sequences divergence promote our understanding of the interspecific diversity of this genus, also provide reference for the rational classification of unsolved species in the future.

Background

The *Pilea* plants are a perennial herb from the family Urticaceae, which mainly distributed in tropical and subtropical, some species distributed in warm temperate regions as well. *Pilea* is a species-rich genus, which is the largest one in the family Urticaceae, and also a larger genus among angiosperms [1]. The leaves of many species in *Pilea* have color spots, which can be used for garden cultivation with ornamental purposes (e.g. *P. cadierei* and *P. mollis*). They are often the main plant groups in the shady
and humid environment of the garden landscape. On the other hand, in Chinese traditional pharmacopeia, several species are recorded as medicinal plants from which a variety of pharmacologically active substances can be extracted [2–4]. For example, *P. peperomioides* is recorded in “Dai medicine” for anti-inflammatory and detoxifying, also used for erysipelas and bone setting. However, it's a group that gets little attention, and there are also little reports about *Pilea*. According to the pharmacopeia, different species have different pharmacological ingredients and medicinal values. Moreover, as many species are similar in morphology, species and population accurate identifications are particularly important for the rational usage of these medicinal plants based on molecular markers.

The genus *Pilea* is also a controversial group in traditional taxonomy, previous studies suggesting that *Sarcopilea* also belongs to this genus [5]. With little attention from experts and scholars, many species being ignored or unresolved, it is difficult for the revision of this species-rich genus. Moreover, some new species have been reported in recent years [6, 7]. It is necessary for us to further study the phylogenetic status of *Pilea* plants. However, relatively little research has been reported on this genus, especially in the field of molecular biology and genomics. Though some researchers have used molecular methods to explore phylogenetic relationships within the genus *Pilea* [1] and its phylogenetic position in the family Urticaceae [5], the selected DNA fragments are one-sided and partial with low bootstrap supports values, which has certain limitations.

The chloroplast is a kind of organelle involved in photosynthesis [8] and energy transformation in plants and algae [9, 10]. Chloroplast genome (referred to as cp genome in the following text) encodes many key proteins that play essential roles in photosynthesis and other metabolic properties [11]. The cp genome sequences have unique characteristics, such as monophyletic inheritance [12], conserved coding region sequences, and a typical quartile structure of genomes. Due to its conservative genome structure and contents, the cp genomes has become an ideal model to resolve plant phylogenies and evaluate biodiversity [13]. Although the cp genomes are relative conserved compared to the nuclear genomes and mitochondrial genomes [14], it also contains highly variable regions that are widely used as molecular markers [15–18]. Chloroplast genomes can provide many molecular markers for species identification in plants, such as chloroplast simple sequence repeats (SSRs), exhibiting great potential in species identification [19, 20]. Unfortunately, no chloroplast genomes of *Pilea* plants have been reported so far.

Here, we first sequenced, assembled and analyzed the cp genomes of four *Pilea* plants. They are common ornamental or medicinal plants (*P. mollis, P. glauca, P. peperomioides* and *P. serpyllacea*). In this study, our main works are as follows: (1) we first sequenced and assembled cp genomes of *Pilea* plants; (2) we analyzed the structural characteristics and sequence divergence of cp genomes in *Pilea*; (3) we identified SSR loci and repeat sequences for further studies on population genetic structure; (4) we inferred the phylogenetic status of *Pilea* in Urticaceae based on the complete cp genome sequences; and (5) we identified the hypervariable regions which could be used as DNA barcodes for identification of this genus.

**Results**
General features of cp genomes

Using Illumina Hiseq sequencing platforms, 5.38–5.89 G clean data from each *Pilea* species were obtained, with the number of clean reads are ranged from 17,935,118 to 19,627,967 (Additional File 1: Table S1). The chloroplast was then assembled based on these data. The 4 cp genomes of *Pilea* are characterized by a typical circular DNA molecule with the length ranging from 150,398 – 152,327 bp. They all have a conservative quartile structure, which is composed of a LSC region (82,063–83,292 bp), a SSC region (17,487 – 18,363 bp) and a pair of IR regions (25,180 – 25,356 bp) (Table 1). The lengths of cp genomes are conserved in this genus. The GC content analysis showed that the overall GC contents ranged from 36.35–36.69% in the 4 cp genomes. Note that the GC contents in IR regions (42.56% – 42.73%) are significantly higher than that in LSC (33.87% – 34.36%) and SSC regions (30.01% – 30.81%). The four cp genomes have been deposited to NCBI (Accession number: MT726015 to MT726018).

| Species | P. glauca | P. mollis | P. peperomioides | P. serpyllacea |
|---------|-----------|-----------|-----------------|---------------|
| Accession number | MT726015 | MT726018 | MT726016 | MT726017 |
| Length (bp) | | | | |
| Total | 151,210 | 150,587 | 152,327 | 150,398 |
| LSC | 82,662 | 82,063 | 83,292 | 82,551 |
| SSC | 17,836 | 17,864 | 18,363 | 17,487 |
| IR | 25,356 | 25,330 | 25,336 | 25,180 |
| GC content (%) | | | | |
| Total | 36.69 | 36.72 | 36.35 | 36.41 |
| LSC | 34.31 | 34.36 | 33.87 | 33.96 |
| IR | 42.64 | 42.65 | 42.73 | 42.56 |
| SSC | 30.81 | 30.76 | 30.01 | 30.23 |
| Gene numbers | | | | |
| Total | 133 | 133 | 133 | 133 |
| Protein-coding gene | 88 | 88 | 88 | 88 |
| tRNA gene | 37 | 37 | 37 | 37 |
| rRNA gene | 8 | 8 | 8 | 8 |

Genome annotation

The cp genomes of four *Pilea* species all comprises 133 genes, among which, 113 are unique genes, including 79 protein-coding genes, 4 rRNA genes and 30 tRNA genes (Additional File 1: Table S2). The gene order and gene numbers of these four species are highly similar, showed conserved genomic
structures. Figure 1 shows the schematic diagram of the cp genomes of *Pilea*. Introns play a significant role in selective gene splicing [21]. Among the 79 protein coding genes annotated, nine unique genes (*rps16, rpoC1, atpF, petB, petD, rpl16, rpl2, ndhB, ndhA*) contain one intron and two unique genes (*ycf3, clpP*) contain two introns. Six unique tRNA genes (*trnK-UUU, trnG-UCC, trnL-UAA, trnV-UAC, trnI-GAU, trnA-UGC*) contain one intron. There are seven protein-coding genes, four rRNA genes, and seven tRNA genes completely duplicated in the IR regions, so they have two copies. The gene *rps12* is a trans-spliced gene, the 5’ end is located in LSC region. However, the 3’ end was found in the IRa and IRb region. These results are similar to other species in the nettle family [22].

**Repeats analysis**

Simple sequence repeats (SSRs), also referred to as the microsatellite sequences, provide a large amount of genetic information [23–25]. Because of its high genetic polymorphism, SSRs are often used for the development of molecular markers and play an important role in the identification of species [26, 27]. In this study, we detected 68, 75, 71, 80 SSRs in the 4 analyzed species, respectively (Fig. 2A, Additional File 1: Table S3). Most SSRs are mononucleotide homopolymers, particularly for A/T, which accounts for 70.75% of the total. Hexanucleotide repeats are detected only in *P. mollis*. These SSR showed high polymorphism, suggesting great potential in the identification of *Pilea* species.

In the cp genomes of *Pilea* species, we detected four types of interspersed repeats. Most of them are forward repeats and palindromic repeats (Fig. 2B). By contrast, there are only two reverse repeats and one complement repeats. The only complement repeats were found in *P. peperomioides*. The detailed sequences showed in Additional File 1: Table S4. Moreover, the length of these short interspersed repeats mainly ranged from 30 to 34 bp. We note that one forward repeats with a length of 102 bp, (detected in *P. serpyllacea*, R13). These longer interspersed repeats thought to be essential for promoting cp genome rearrangements [28, 29]. Whether or not these repeats have caused the rearrangement of the cp genomes of *Pilea* species are interesting questions.

**Contraction and expansion analysis of IR regions**

The contraction and expansion of IR regions are considered to be an important reason for the length diversity in cp genomes [30]. Besides, with the expanded/contracted of the IR regions, genes near the border have an opportunity access to IR or single-copy regions [31]. We retrieved the published cp genomes of six species from Urticaceae, and compared them with the four *Pilea* species. We observed several genes span or near the boundary of IR and single copy regions. They mainly are *rps19, rpl22, rpl2, ycf1, ndhF* and *trnH* (Fig. 3). It’s worth noting that, an abnormal expansion of IR regions was observed in *Gonostegia hirta*. The IR regions are over 30,000 bp in *G. hirta* and more genes access to the IR regions (e.g. *rpl36* and *rps19*). However, the length of IR regions in the other nine species are about 25,000 bp, and *rps19* gene span the LSC/IRb boundary except for *D. iners* and *H. tenella*. The former *rps19* gene is in LSC region, while the latter is completely in IR regions. Besides, *trnH* gene completely accesses to IR regions in *H. tenella*, obtained two copies. It can be seen that the genomic structure, gene order and numbers of some species in Urticaceae have changed obviously.
Furthermore, gene \( ycf1 \) crosses the SSC/IRa boundary, most of which located in the SSC region. The length of \( ycf1 \) gene in the four \( Pilea \) species varies widely, indicating the possibility of sequence differences. Surprisingly, we annotated two copies of \( ycf1 \) in the four \( Pilea \) plants, they cross IRb/SSC boundary and are not annotated in other species. Sequence alignment found that the two copies of \( ycf1 \) are existed in other taxa, indicating that previous annotation are imperfect, although one of the two copies is a fragment of \( ycf1 \), and are generally considered to be a pseudogene. Interestingly, except for \( E. dissectum \), a small fraction of the \( ndhF \) gene (less than 100 bp) crosses the IRb/SSC regions, which means that fragments of \( ycf1 \) have an overlap with \( ndhF \) in \( Pilea \) species. The overlapping areas are 108 bp. These results are also observed in \( Arabidopsis \), the overlaps are about 30 bp [32]. Whether or not these overlaps affect the transcription or translation of these proteins is also an interesting subject.

**Genome divergence**

To evaluate genomic divergence, a sequence identity analysis based on mVISTA [33, 34] was performed between 4 \( Pilea \) species, with the reference being the cp genome of \( P. peperomioides \). We observed varying degrees of sequence divergence, especially in LSC and SSC regions. In contrast, the IR regions are more conservative. Most of these highly variable regions are observed in Conserved Non-Coding Sequences (CNS) (Fig. 4). However, the regions with the greatest sequence divergence are found in protein-coding regions, which is gene \( ycf1 \). The coding regions of \( ycf1 \) in the four \( Pilea \) species showed significant differences, and the similarity are even less than 50% in some fragments. Overall, the analyzed genomic sequences showed high levels of sequence divergence, suggesting a high level of genetic diversity in the genus \( Pilea \).

To quantify the levels of DNA polymorphism, the 4 cp genomes were aligned and analyzed by using DnaSP v6.0 [35]. We detected 8 hypervariable regions with Pi value exceed 0.06 (Fig. 5), including \( petN-psbM \) (Pi = 0.06067); \( psbZ-trnG-GCC \) (Pi = 0.07067); \( trnT-UGU-trnL-UAA \) (Pi = 0.06433); \( accD-psbl \) (Pi = 0.06003); \( ndhF-rpl32 \) (Pi = 0.06100); \( rpl32-trnL-UAG \) (Pi = 0.06800); \( ndhA-intron \) (Pi = 0.06533) and the most regions of gene \( ycf1 \) (Pi values are ranged from 0.07367 to 0.17067). The Pi values are listed in the parentheses. It's noteworthy that most regions of the cp genome sequences had Pi values more than 0.02 (except for IR regions), exhibiting abundant polymorphism of cp genome sequences.

**Nucleotide variations in protein-coding genes**

The protein-coding regions are highly conserved in cp genomes [36]. We analyzed the protein-coding sequences of 79 identified unique orthologous genes in 4 \( Pilea \) taxa. Surprisingly, these protein-coding genes also showed high levels of variation (Fig. 6A, Additional File 1: Table S5). Of the 79 shared genes, 63 had a mutation rate of more than 2%, and 30 had a mutation rate of more than 4%. The gene with the highest mutation rates is \( ycf1 \) (16.62%), then followed by \( matK \) (10.54%), \( ccsA \) (8.74%) and \( rps15 \) (8.42%). Only two genes (\( psbJ \) and \( psbL \)) showed extreme conservation without any variable sites. Moreover, we observed a total of 11 genes (\( ycf1, ndhF, rps19, accD, rpoC2, rps16, rpoA, rpl20, ndhD, rpoC1 \) and \( ycf2 \)) with InDels in nucleotide sequences by using DNAsP 6.0 [35]. Among which, \( ycf1 \) had 35 InDels, then followed by \( ycf2 \) (9), \( accD \) (4) and \( rpoC2 \) (3). Considering that the protein-coding regions are
highly conserved, protein-coding sequences with high nucleotide mutation rates are usually infrequent in the same genus, these results indicated that the *Pilea* plants had high genetic diversity and there are great differences among species.

In this study, synonymous (dS) and nonsynonymous (dN) substitution rates, along with dN/dS were estimated for the 79 shared genes in parallel by using PAML v4.9 [37]. Among the 79 genes, *ycf1*, *matK*, *ccsA* and *rps15* had relatively higher dN values, and *rps16*, *rpl32*, *ndhF* and *psaJ* had higher dS values (Fig. 6B, Additional file 1: Table S6). Most genes exhibited considerably low values of dN/dS (less than 0.6), implying most of the protein-coding genes were under purifying selection during evolution. However, the dN/dS ratio of three genes (*rpl36*, *clpP* and *accD*) was between 0.6 and 1.0. Moreover, the dN/dS ratio was greater than 1.0 for *petL*, *rps12*, *ycf1* and *ycf2*, indicating that they have been under positive selection during evolution. This results clearly indicate that cp genes in different plant species of *Pilea* may be subjected to diverse selection pressures.

**Phylogenetic analysis**

Compared to nuclear and mitochondrial genomes, the cp genomes are highly conserved, and it has been widely used in phylogenetic and evolutionary studies [38–40]. With the development of high throughput sequencing technology, the chloroplast genome sequence plays an important role in species identification as a super barcode [41, 42]. In this study, we constructed the maximum likelihood (ML) trees by using the complete cp genome sequences as the data sets (detailed materials are shown in Additional File 1: Table S7). The phylogenetic tree has high bootstrap supports in all nodes, shows the reliability of the phylogeny recovered (Fig. 7).

Our phylogenetic tree displayed two clades clearly, and then further diversified into four subclades with 100% bootstrap supports (ML). These four subclades correspond to four subfamilies, they are Boehmerioideae, Cecropioideae, Lecanthoideae and Urticoideae, respectively. This is consistent with the traditional classification [5]. All the 4 *Pilea* species clustered together (all nodes have BS = 100 for ML), and form a monophyletic group, which is a sister group to *Elatostema*. They all are belonging to subfamily Lecanthoideae.

**Discussion**

**Conserved genome structure and gene content**

We first reported the cp genome sequences of four *Pilea* species. Our assembly results showed that the length of 4 cp genomes ranged from 150,398 bp to 152,327 bp. This result is similar to most Urticaceae plants [43, 44]. In our study, the longest and shortest cp genome sequences in Urticaceae are 159,085 bp (*Gonostegia Hirta*) and 146,842 bp (*Hesperocnide Tenella*), respectively. This suggests that the cp genomes of Urticaceae may have undergone different evolutionary processes. Among our four *Pilea* taxa, the longest genome sequence is *P. peperomioides* (152,327 bp) and the shortest is *P. serpyllacea*
Structurally, they are highly similar to most angiosperms, and we didn’t detect gene gain/loss, suggesting that cp genomes are still relatively conserved.

In this study, we detected SSRs and repeat sequences in the four cp genomes. Of the total 294 SSRs, 215 are mononucleotide repeats, accounting for the majority of all SSRs (73.13%). These mononucleotide repeats are mainly A/T repeats, and they have a significant impact on the overall G/C content of the genomes [45, 46]. These SSR sequences are often composed of simple repeating units such as polyadenine (PolyA) or polythymine (Poly T) repeats. With the length polymorphism in different species, they are often used as molecular markers. There are abundant SSR loci in cp genomes, which have been applied in species identification [47, 48].

The variation of IR regions is a common phenomenon in angiosperms. Compared with the overall absence of one IR region [49–51], the expansion/contraction of IR regions are more common in angiosperms [52, 53]. By comparative analysis, we found that G. hirta significantly expanded the IR regions, which also led to an increase in the overall length of the cp genomes. In our four Pilea plants, the length of IR regions ranged from 25,180 bp to 25,356 bp, showing no significant difference. As far as the boundary regions of IR/SC is concerned, the position of genes near the boundary in the four Pilea species is similar to most angiosperms. This indicates that the Pilea plants did not experience significant expansion/contraction in the IR regions. However, we observed the overlaps of ycf1 and SSC region (4,634 bp) in P. peperomioides was longer than other three species (4,203 bp – 4,314 bp), and the overlaps with IRa are similar (803 bp – 843 bp). This suggests that there is a significant difference in ycf1 gene sequences.

**High levels of sequences divergence reveals interspecific diversity in Pilea**

In our chloroplast comparative genomics analysis, we compared the whole cp genome sequences based on mVISTA. Specifically, we also calculated the percentage of variable sites and estimated the ratios of dN/dS rates among 79 orthologous protein-coding genes. Similar to most angiosperms, the non-coding regions of Pilea plants showed higher polymorphism than coding regions. Surprisingly, we also found high levels of sequence differences in the coding regions of Pilea plants. Of the 79 orthologous genes identified, 63 had a mutation rate of more than 2%, and 30 had a mutation rate of more than 4%. This is rare in other genera, because usually only the ycf1 gene had a high mutation rate [54]. The mutation rate of ycf1 gene in the four Pilea plants is an astonishing 16.62%. Also, a total of 35 InDels were detected, including a large fragment insertion in P. peperomioides (177 bp, data not shown). These InDels caused an increase in the length of the ycf1 gene in P. peperomioides. In addition, unusually high nucleotide mutation rates also observed in matK, ccsA and other genes.

In general, dN changes are subject to bidirectional effects of varied mutation rates and selective constraints. The ratio of dN/dS greater than 1 is thought to be a sign that the gene has experienced selection pressures. In our study, the dN/dS ratio indicates that four genes (petL, rps12, ycf1 and ycf2) may have undergone positive selection. The rapid evolution of protein-coding genes is closely related to
the adaptive evolution of species [55, 56], indicating that Pilea plants may have experienced a rapid evolutionary process, result in species-rich of Pilea.

Although the sequence divergence depends on the species we are comparing, such high levels of sequence differences exhibited a high level of genetic diversity in Pilea taxa. Previous estimates for the species number in Pilea are varied widely. The Flora of China puts the number at about 400 [57]. However, some scholars believe that there are more than 700 species [58]. How many species in this genus is still unclear so far. To date, we queried 868 results on The Plant List (http://www.plantlist.org/), including 286 accepted species, 163 Synonym and 419 unresolved species. A large number of unresolved species hints at complex interspecific relationships. Although the number of species members of Pilea are not definitively determined, our research supports that this is a rapidly evolving group, and conservative estimates should be deliberation.

**8 hypervariable regions could be used as a potential DNA barcodes**

Meanwhile, we used DNASP v6.0 to quantify DNA sequence polymorphism by conducting a sliding window analysis (window length: 500 bp, Step size: 500 bp). Similar to the results of mVISTA, most regions have high Pi values except IR regions, which means that several regions have the potential to develop molecular markers. We recommend eight hypervariable regions, including petN-psbM (Pi = 0.06067); psbZ-trnG-GCC (Pi = 0.07067); trnT-UGU-trnL-UAA (Pi = 0.06433); accD-psbl (Pi = 0.06003); ndhF-rpl32 (Pi = 0.06100); rpl32-trnL-UAG (Pi = 0.06800); ndhA-intron (Pi = 0.06533) and almost the entire ycf1 gene (Pi values are ranged from 0.07367 to 0.17067), as a potential molecular marker for Pilea plants. In particular, the gene ycf1 with a large number of InDels can be used as specific molecular markers, which is of great significance for us to correctly identify and rationally utilize the medicinal taxa from this genus.

**Phylogenetic status of Pilea based on cp genome sequences**

Moreover, the phylogenetic status of Pilea in Urticaceae was analyzed based on the complete cp genomic sequences. In a one-sided analysis based on chloroplast genome sequences, Pilea and Elatostema are sister groups of each other, both belonging to the subfamily Lecanthoideae. This is consistent with the results of traditional classification studies [5]. However, due to the matrilineal inheritance of the chloroplast genomes [59], the results are restricted. Accurate phylogenetic relationships need a comprehensive analysis of nuclear and organelle genes [60]. Furthermore, the relationships between Pilea and other plants of family Urticaceae need more cp genome sequencing in the future.

**Conclusions**

In this study, four cp genomes of Pilea plants were sequenced, assembled and annotated, which was the first report in this genus. By comparison, we found the cp genomes of four species have similar structural characteristics, and a typical quartile structure similar to that of most angiosperms. Unusually, the
genome sequences of four species, including the relatively conserved protein-coding regions, showed high levels of divergence. We glimpsed a rich genetic diversity in *Pilea*. The interspecific diversity of *Pilea* may be more abundant than we previously thought. This is of reference significance for us to evaluate the number of species in this genus. In summary, we provide 4 high quality chloroplast reference genomes, and the results obtained here provide valuable information for the understanding of the genetic diversity and contribute to the resource utilization of *Pilea* plants in the future.

**Methods**

**Plant material, DNA extraction and Sequencing**

The fresh leaves of four *Pilea* plants were collected from Guangzhou, Kunming and Suqian, respectively. All the samples were saved at the Herbarium of Southwest University, Chongqing, China. The detailed information for the plant samples shown in Additional File 1: Table S8. The total genomic DNA was extracted by using CTAB method [61]. The DNA library with an insert size of 350 bp was constructed using the NEBNext® library building kit [62] and sequenced by using the Hiseq Xten PE150 sequencing platform. Sequencing produced a total of 5.4–5.9 G raw data per species. Clean data were obtained by removing low-quality sequences: sequences with a quality value of Q < 19 accounted for more than 50% of the total base, and sequences with more than 5% bases being “N”.

**Genome assembly and annotation**

The de novo genome assembly from the clean data was accomplished utilizing the NOVOPlasty (v.2.7.2) [63] with the k-mer length of 39 bp and a sequence fragment of *rbcL* gene from maize as the seed sequence. The correctness of the assembly was confirmed by using Bowtie2 (v2. 0.1) [64] to manually edit and map all raw reads to the assembled genome sequence under the default settings. The cp genome was annotated initially by using CPGAVAS2 [65] using the reference genome (*Elatostema dissectum*, GenBank: NC_047192.1). Geseq was then used to confirm the annotation results [66]. Furthermore, the annotations with problems were manually edited by using Apollo [67]. The genome maps were drawn by OGDRAW [68]. The genome sequence has been deposited in GenBank with accession numbers: MT726015 to MT726018.

**Repeats and SSR analysis**

The GC content was conducted by using the cusp program provided by EMBOSS (v6.3.1) [69]. The simple sequence repeats (SSRs) were identified using the Online website MISA (https://webblast.ipk-gatersleben.de/misa/), including mono-, di-, tri-, tetra-, penta-, and hexanucleotides with the minimum numbers were 10, 5, 4, 3, 3, and 3, respectively [70]. Additionally, REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) was used to calculate palindromic repeats, forward repeats, reverse repeats, and complement repeats with the settings: Hamming Distance was three, and Minimal Repeat Size was 30 bp [71].

**Genome comparison**
The cp genomes of 4 *Pilea* species, were compared by using shuffle-LAGAN mode in mVISTA [33, 34] to identify interspecific variations (http://genome.lbl.gov/vista/mvista/submit.shtml). A total of 79 orthologous genes among the 4 species were identified and extracted by using Phyllosuite [72]. The corresponding nucleotide sequences were aligned by using MAFFT (v 7.450) [73] implemented in Phyllosuite. We used MEGA v6.0 [74] to calculate the percentage of variable sites in protein-coding genes. We conducted a sliding window analysis (window length: 500 bp, Step size: 500 bp) by using DnaSP v6.0 [35] to calculate the nucleotide polymorphism (Pi) among the 4 species. Lastly, IRscope (https://irscope.shinyapps.io/irapp/) was used for visualizing the IR boundaries in these cp genomes [75].

**Analysis of nucleotide substitution rate**

The protein-coding sequences in the previous step were processed in parallel. We used the CODEML module in PAML v.4.9 [37] to estimate rates of nucleotide substitution, including dN (nonsynonymous), dS (synonymous), and the ratio of nonsynonymous to synonymous rates (dN/dS). The detailed parameters were: CodonFreq = 2 (F3 × 4 model); model = 0 (allowing a single dN/dS value to vary among branches); cleandata = 1 (remove sites with ambiguity data); other parameters in the CODEML control file were left at default settings. The phylogeny tree structure of each genes were generated by using the Maximum Likelihood (ML) method implemented in RaxML (v8.2.4) [76].

**Phylogenetic Analysis**

The cp genome sequences of 19 species belonging to the family Urticaceae, were downloaded from GenBank (NCBI, https://www.ncbi.nlm.nih.gov/). These species belonging to 4 subfamilies (Additional File 1: Table S7). Two species, *Morus indica* (Moraceae) and *Ficus carica* (Moraceae), were used as outgroups. The complete cp genome sequences were aligned by using MAFFT (https://mafft.cbrc.jp/alignment/server/) online version 7.471 [73]. These aligned sequences were used to construct the phylogenetic trees by using the Maximum Likelihood (ML) method implemented in RaxML (v8.2.4) [76]. The parameters were “raxmlHPC-PTHREADS-SSE3 -f a -N 1000 -m GTRGAMMA -x 551314260 -p 551314260”. The bootstrap analysis was performed with 1,000 replicates.

**Abbreviations**

Cp: Chloroplast; SSR: Simple sequence repeat; CNS: Conserved Non-Coding Sequences; IRs: Inverted repeats; LSC: Large single-copy; SSC: Small single-copy; ML: Maximum-likelihood; BS: Branch support; PolyA: polyadenine; PolyT: polythymine; dS: synonymous substitution rates, dN: nonsynonymous substitution rates; DnaSP: DNA Sequences Polymorphism; CTAB: Cetyl trimethylammonium bromide; NCBI: National Center for Biotechnology Information; Pi: Nucleotide diversity (π).

**Declarations**

**Availability of data and materials**
The annotated chloroplast genome sequences of four *Pilea* plants were deposited in GeneBank (https://www.ncbi.nlm.nih.gov/) with Accession number: MT726015 to MT726018. All the samples are saved at the Herbarium of Southwest University, Chongqing, China. All other data and material generated in this manuscript are available from the corresponding author upon reasonable request.

**Ethics approval and consent to participate**

The four collected *Pilea* species are widely distributed in China as ornamental plants. Experimental researches do not include the genetic transformation, preserving the genetic background of the species used, and any other processes requiring ethics approval.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests

**Funding**

This work was supported by the National Natural Science Foundation of China [31772260] and Chongqing Study Abroad Innovation Project [cx2019052]. The funders were not involved in the study design, data collection, and analysis, decision to publish, or manuscript preparation.

**Author Contributions**

JieY conceived the study and designed experiments; JingY collected the samples and extracted DNA for sequencing by using the Illumina platform; FH assembled and annotated the cp genomes; JMT, SYZ and JLL carried out the comparative chloroplast analysis; JLL drafted the manuscript. All authors have read and approved the final manuscript.

**Acknowledgements**

The authors are grateful to the technical support provided by Novogene (Tianjin) and Professor Chang Liu.

**References**

1. Monro AK. The revision of species-rich genera: a phylogenetic framework for the strategic revision of *Pilea* (Urticaceae) based on cpDNA, nrDNA, and morphology. Am J Bot. 2006;93(3):426–41.
2. Zhou Y, Li LY, Ren HC, Qin RD, Li Q, Tu PF, Dou GF, Zhang QY, Liang H. Chemical constituents from the whole plants of Pilea cavaleriei Levl subsp. cavaleriei. Fitoterapia. 2017;119:100–7.
doi:10.1016/j.fitote.2017.04.010.

3. Prabhakar KR, Veerapur VP, Bansal P, Parihar VK, Reddy Kandadi M, Bhagath Kumar P, Priyadarsini KI, Unnikrishnan MK. Antioxidant and radioprotective effect of the active fraction of Pilea microphylla (L.) ethanolic extract. Chemico-biological interactions. 2007; 165(1):22–32.
doi:10.1016/j.cbi.2006.10.007.

4. Modarresi Chahardehi A, Ibrahim D, Fariza Sulaiman S, Antioxidant, Antimicrobial Activity and Toxicity Test of Pilea microphylla. International journal of microbiology. 2010; 2010:826830.
doi:10.1155/2010/826830.

5. Wu ZY, Monro AK, Milne RI, Wang H, Yi TS, Liu J, Li DZ. Molecular phylogeny of the nettle family (Urticaceae) inferred from multiple loci of three genomes and extensive generic sampling. Mol Phylogenet Evol. 2013;69(3):814–27. doi:10.1016/j.ympev.2013.06.022.

6. Dorr LJ, Stergios B. Four new species of Andean Pilea (Urticaceae), with additional notes on the genus in Venezuela. PhytoKeys. 2014(42):57–76. doi:10.3897/phytokes.42.8455.

7. Monro AK, Wei YG, Chen CJ. Three new species of Pilea (Urticaceae) from limestone karst in China. PhytoKeys. 2012(19):51–66. doi:10.3897/phytokes.19.3968.

8. Szabò I, Spetea C. Impact of the ion transportome of chloroplasts on the optimization of photosynthesis. J Exp Bot. 2017;68(12):3115–28. doi:10.1093/jxb/erx063.

9. Mullineaux PM, Exposito-Rodriguez M, Laissue PP, Smirnoff N. ROS-dependent signalling pathways in plants and algae exposed to high light: Comparisons with other eukaryotes. Free Radic Biol Med. 2018;122:52–64. doi:10.1016/j.freeradbiomed.2018.01.033.

10. Pollari M, Ruotsalainen V, Rantamaki S, Tyystjarvi E, Tyystjarvi T. Simultaneous inactivation of sigma factors B and D interferes with light acclimation of the cyanobacterium Synechocystis sp. strain PCC 6803. J Bacteriol. 2009;191(12):3992–4001. doi:10.1128/JB.00132-09.

11. Wang Z, Zhu XG, Chen Y, Li Y, Hou J, Li Y, Liu L. Exploring photosynthesis evolution by comparative analysis of metabolic networks between chloroplasts and photosynthetic bacteria. BMC Genom. 2006;7:100. doi:10.1186/1471-2164-7-100.

12. Brandrud MK, Baar J, Lorenzo MT, Athanasiadis A, Bateman RM, Chase MW, Hedren M, Paun O. Phylogenomic Relationships of Diploids and the Origins of Allopolyploids in Dactylorhiza (Orchidaceae). Syst Biol. 2020;69(1):91–109. doi:10.1093/sysbio/syz035.

13. Shin DH, Lee JH, Kang SH, Ahn BO, Kim CK. The Complete Chloroplast Genome of the Hare's Ear Root, Bupleurum falcatum: Its Molecular Features. Genes (Basel) 2016; 7(5).
doi:10.3390/genes7050020.

14. Koch L. Genetic variation: Nuclear and mitochondrial genome interplay. Nature reviews Genetics. 2016;17(9):502. doi:10.1038/nrg.2016.96.
15. Liu ML, Fan WB, Wang N, Dong PB, Zhang TT, Yue M, Li ZH. Evolutionary Analysis of Plastid Genomes of Seven Lonicera L. Species: Implications for Sequence Divergence and Phylogenetic Relationships. Int J Mol Sci 2018; 19(12). doi:10.3390/ijms19124039.

16. Liu X, Zhou B, Yang H, Li Y, Yang Q, Lu Y, Gao Y. Sequencing and Analysis of Chrysanthemum carinatum Schousb and Kalimeris indica. The Complete Chloroplast Genomes Reveal Two Inversions and rbcL as Barcoding of the Vegetable. Molecules (Basel, Switzerland). 2018; 23(6). doi:10.3390/molecules23061358.

17. Pang X, Liu H, Wu S, Yuan Y, Li H, Dong J, Liu Z, An C, Su Z, Li B. Species Identification of Oaks (Quercus L., Fagaceae) from Gene to Genome. Int J Mol Sci 2019; 20(23). doi:10.3390/ijms20235940.

18. Thakur VV, Tiwari S, Tripathi N, Tiwari G. Molecular identification of medicinal plants with amplicon length polymorphism using universal DNA barcodes of the atpF-atpH, trnL and trnH-psbA regions. 3 Biotech. 2019;9(5):188. doi:10.1007/s13205-019-1724-6.

19. Ebert D, Peakall R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol Ecol Resour. 2009;9(3):673–90. doi:10.1111/j.1755-0998.2008.02319.x.

20. Huang LS, Sun YQ, Jin Y, Gao Q, Hu XG, Gao FL, Yang XL, Zhu JJ, El-Kassaby YA, Mao JF. Development of high transferability cpSSR markers for individual identification and genetic investigation in Cupressaceae species. Ecol Evol. 2018;8(10):4967–77. doi:10.1002/ece3.4053.

21. Plangger R, Juen MA, Hoernes TP, Nußbaumer F, Kremser J, Strebitzer E, Klingler D, Erharter K, Tollinger M, Erlacher MD, et al. Branch site bulge conformations in domain 6 determine functional sugar puckers in group II intron splicing. Nucleic Acids Res. 2019;47(21):11430–40. doi:10.1093/nar/gkz965.

22. Kim S-C, Baek S-H, Hong K-N, Lee J-W. Characterization of the complete chloroplast genome of Koelreuteria paniculata (Sapindaceae). Conservation Genetics Resources. 2017; 10. doi:10.1007/s12686-017-0767-4.

23. Ma SJ, Sa KJ, Hong TK, Lee JK. Genetic diversity and population structure analysis in Perilla crop and their weedy types from northern and southern areas of China based on simple sequence repeat (SSRs). Genes Genomics. 2019;41(3):267–81. doi:10.1007/s13258-018-0756-3.

24. Seyoum M, Du XM, He SP, Jia YH, Pan Z, Sun JL. Analysis of genetic diversity and population structure in upland cotton (Gossypium hirsutum L.) germplasm using simple sequence repeats. J Genet. 2018;97(2):513–22.

25. Yang X, Xu Y, Shah T, Li H, Han Z, Li J, Yan J. Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica. 2011;139(8):1045–54. doi:10.1007/s10709-011-9606-9.

26. Guang XM, Xia JQ, Lin JQ, Yu J, Wan QH, Fang SG. IDSSR: An Efficient Pipeline for Identifying Polymorphic Microsatellites from a Single Genome Sequence. Int J Mol Sci 2019; 20(14). doi:10.3390/ijms20143497.
27. Guo Q, Li X, Yang S, Yang Z, Sun Y, Zhang J, Cao S, Dong L, Uddin S, Li Y. Evaluation of the Genetic Diversity and Differentiation of Black Locust (Robinia pseudoacacia L.) Based on Genomic and Expressed Sequence Tag-Simple Sequence Repeats. Int J Mol Sci 2018; 19(9). doi:10.3390/ijms19092492.

28. Lee HO, Joh HJ, Kim K, Lee SC, Kim NH, Park JY, Park HS, Park MS, Kim S, Kwak M, et al. Dynamic Chloroplast Genome Rearrangement and DNA Barcoding for Three Apiaceae Species Known as the Medicinal Herb "Bang-Poong". Int J Mol Sci 2019; 20(9). doi:10.3390/ijms20092196.

29. McCann J, Jang TS, Macas J, Schneeweiss GM, Matzke NJ, Novak P, Stuessy TF, Villasenor JL, Weiss-Schneeweiss H. Dating the Species Network: Allopolyploidy and Repetitive DNA Evolution in American Daisies (Melampodium sect. Melampodium, Asteraceae). Syst Biol. 2018;67(6):1010–24. doi:10.1093/sysbio/syy024.

30. Goulding SE, Olmstead RG, Morden CW, Wolfe KH. Ebb and ow of the chloroplast inverted repeat. Mol Gen Genet. 1996;252(1–2):195–206. doi:10.1007/bf02173220.

31. Wang W, Chen S, Zhang X. Whole-Genome Comparison Reveals Divergent IR Borders and Mutation Hotspots in Chloroplast Genomes of Herbaceous Bamboos (Bambusoideae: Olyreae). Molecules (Basel, Switzerland). 2018; 23(7). doi:10.3390/molecules23071537.

32. Asaf S, Khan AL, Khan MA, Waqas M, Kang SM, Yun BW, Lee IJ. Chloroplast genomes of Arabidopsis halleri ssp. gemmifera and Arabidopsis lyrata ssp. petraea: Structures and comparative analysis. Sci Rep. 2017;7(1):7556. doi:10.1038/s41598-017-07891-5.

33. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004; 32(Web Server issue):W273–9. doi:10.1093/nar/gkh458.

34. Thiel T, Michalek W, Varshney RK, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003;106(3):411–22. doi:10.1007/s00122-002-1031-0.

35. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets
10.1093/molbev/msx248
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Molecular biology evolution. 2017; 34(12):3299–302. doi:10.1093/molbev/msx248.

36. Hong Z, Wu Z, Zhao K, Yang Z, Zhang N, Guo J, Tembrock LR, Xu D. Comparative Analyses of Five Complete Chloroplast Genomes from the Genus Pterocarpus (Fabaceae). Int J Mol Sci 2020; 21(11). doi:10.3390/ijms21113758.

37. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology evolution. 2007;24(8):1586–91. doi:10.1093/molbev/msm088.

38. Du YP, Bi Y, Yang FP, Zhang MF, Chen XQ, Xue J, Zhang XH. Complete chloroplast genome sequences of Lilium: insights into evolutionary dynamics and phylogenetic analyses. Scientific reports. 2017;7(1):5751. doi:10.1038/s41598-017-06210-2.
39. Guo S, Guo L, Zhao W, Xu J, Li Y, Zhang X, Shen X, Wu M, Hou X. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of Paeonia ostii. Molecules. 2018; 23(2). doi:10.3390/molecules23020246.

40. Shen X, Guo S, Yin Y, Zhang J, Yin X, Liang C, Wang Z, Huang B, Liu Y, Xiao S, et al. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of Aster tataricus. Molecules. 2018; 23(10). doi:10.3390/molecules23102426.

41. Fu CN, Wu CS, Ye LJ, Mo ZQ, Liu J, Chang YW, Li DZ, Chaw SM, Gao LM. Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (Taxus) worldwide. Sci Rep. 2019;9(1):2773. doi:10.1038/s41598-019-39161-x.

42. Krawczyk K, Nobis M, Myszczyński K, Klichowska E, Sawicki J. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: Stipa). Sci Rep. 2018;8(1):1924. doi:10.1038/s41598-018-20399-w.

43. Fu L-F, Xin Z-B, Wen F, Li S, Wei Y-G. Complete chloroplast genome sequence of Elatostema dissectum (Urticaceae). Mitochondrial DNA Part B. 2019;4:838–9. doi:10.1080/23802359.2019.1567292.

44. Wang R-N, Liu J, Li Z-H, Wu Z-Y. Complete chloroplast genome sequences of Debregeasia orientalis (Urticaceae). Mitochondrial DNA Part B. 2019;4(1):1830–1. doi:10.1080/23802359.2019.1604186.

45. Gichira AW, Avoga S, Li Z, Hu G, Wang Q, Chen J. Comparative genomics of 11 complete chloroplast genomes of Senecioneae (Asteraceae) species: DNA barcodes and phylogenetics. Bot Stud. 2019;60(1):17. doi:10.1186/s40529-019-0265-y.

46. Li W, Zhang C, Guo X, Liu Q, Wang K. Complete chloroplast genome of Camellia japonica genome structures, comparative and phylogenetic analysis. PloS one. 2019;14(5):e0216645. doi:10.1371/journal.pone.0216645.

47. Jiang M, Chen H, He S, Wang L, Chen AJ, Liu C. Sequencing, Characterization, and Comparative Analyses of the Plastome of Caragana rosea var. rosea. Int J Mol Sci 2018; 19(5). doi:10.3390/ijms19051419.

48. Su Y, Liu Y, Li Z, Yang Z, Zhong M, Zhang Y. QTL Analysis of Head Splitting Resistance in Cabbage (Brassica oleracea L. var. capitata) Using SSR and InDel Makers Based on Whole-Genome Re-Sequencing. PLoS One. 2015;10(9):e0138073. doi:10.1371/journal.pone.0138073.

49. Rousseau-Gueutin M, Bellot S, Martin GE, Boutte J, Chelaifa H, Lima O, Michon-Coudouel S, Naquin D, Salmon A, Ainoouche K, et al. The chloroplast genome of the hexaploid Spartina maritima (Poaceae, Chloridoideae): Comparative analyses and molecular dating. Molecular phylogenetics evolution. 2015;93:5–16. doi:10.1016/j.ympev.2015.06.013.

50. Zheng W, Chen J, Hao Z, Shi J. Comparative Analysis of the Chloroplast Genomic Information of Cunninghamia lanceolata (Lamb.) Hook with Sibling Species from the Genera Cryptomeria D. Don, Taiwania Hayata, and Calocedrus Kurz. Int J Mol Sci 2016; 17(7). doi:10.3390/ijms17071084.

51. Hao Z, Cheng T, Zheng R, Xu H, Zhou Y, Li M, Lu F, Dong Y, Liu X, Chen J, et al. The Complete Chloroplast Genome Sequence of a Relict Conifer Glyptostrobus pensilis: Comparative Analysis and
Insights into Dynamics of Chloroplast Genome Rearrangement in Cupressophytes and Pinaceae. PLoS one. 2016;11(8):e0161809. doi:10.1371/journal.pone.0161809.

52. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. The New phytologist. 2016;209(4):1747–56. doi:10.1111/nph.13743.

53. He J, Yao M, Lyu RD, Lin LL, Liu HJ, Pei LY, Yan SX, Xie L, Cheng J. Structural variation of the complete chloroplast genome and plastid phylogenomics of the genus Asteropyrum (Ranunculaceae). Scientific reports. 2019;9(1):15285. doi:10.1038/s41598-019-51601-2.

54. Dong W, Xu C, Li C, Sun J, ZuO Y, Shi S, Cheng T, Guo J, Zhou S. ycf1, the most promising plastid DNA barcode of land plants. Scientific reports. 2015;5:8348. doi:10.1038/srep08348.

55. Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, Li ZH. Molecular Evolution of Chloroplast Genomes of Orchid Species: Insights into Phylogenetic Relationship and Adaptive Evolution. Int J Mol Sci 2018; 19(3). doi:10.3390/ijms19030716.

56. Huang Y, Wang J, Yang Y, Fan C, Chen J. Phylogenomic Analysis and Dynamic Evolution of Chloroplast Genomes in Salicaceae. Frontiers in plant science. 2017;8:1050. doi:10.3389/fpls.2017.01050.

57. Chen C-j. Flora of China. 5. Beijing: Science Press; St Louis: Missouri Botanical Garden Press.; 2003. pp. 92–120.

58. Monro AK. Three New Species, and Three New Names in Pilea (Urticaceae) from New Guinea. Contributions to the Flora of Mt Jaya XV. Kew Bull. 2004;59(4):573–9. doi:10.2307/4110914.

59. Christie JR, Beekman M. Uniparental Inheritance Promotes Adaptive Evolution in Cytoplasmic Genomes. Molecular biology evolution. 2017;34(3):677–91. doi:10.1093/molbev/msw266.

60. 10.3389/fpls.2016.02022
   Liu X, Wang Z, Shao W, Ye Z, Zhang J. Phylogenetic and Taxonomic Status Analyses of the Abaso Section from Multiple Nuclear Genes and Plastid Fragments Reveal New Insights into the North America Origin of Populus (Salicaceae). Frontiers in plant science. 2016; 7:2022. doi:10.3389/fpls.2016.02022.

61. Arseneau JR, Steeves R, Laflamme M. Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. Mol Ecol Resour. 2017;17(4):686–93. doi:10.1111/1755-0998.12616.

62. Emerman AB, Bowman SK, Barry A, Henig N, Patel KM, Gardner AF, Hendrickson CL. NEBNext Direct: A Novel, Rapid, Hybridization-Based Approach for the Capture and Library Conversion of Genomic Regions of Interest. Curr Protoc Mol Biol 2017; 119:7.30.31–37.30.24. doi:10.1002/cpmb.39.

63. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty. De novo assembly of organelle genomes from whole genome data. Nucleic acids research. 2016. doi:10.1093/nar/gkw955.

64. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25. doi:10.1186/gb-2009-10-3-r25.
56. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic acids research. 2019;47(W1):W65-w73. doi:10.1093/nar/gkz345.
57. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6-w11. doi:10.1093/nar/gkx391.
58. Misra S, Harris N. Using Apollo to browse and edit genome annotations. Curr Protoc Bioinformatics. 2006. doi:10.1002/0471250953.bi0905s12. Chap. 9:Unit 9.5.
59. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic acids research. 2019;47(W1):W59-w64. doi:10.1093/nar/gkz238.
60. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16(6):276–7. doi:10.1016/s0168-9525(00)02024-2.
61. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5. doi:10.1093/bioinformatics/btx198.
62. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic acids research. 2001;29(22):4633–42. doi:10.1093/nar/29.22.4633.
63. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55. doi:10.1111/1755-0998.13096.
64. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic acids research. 2019;47(W1):W5-w10. doi:10.1093/nar/gkz342.
65. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9. doi:10.1093/molbev/mst197.
66. Amiryousefi A, Hyvonen J, Poczai P. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics. 2018;34(17):3030–1. doi:10.1093/bioinformatics/bty220.
67. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies 10.1093/bioinformatics/btu033
68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014; 30(9):1312–3. doi:10.1093/bioinformatics/btu033.

Figures
Figure 1

Graphic representation of features identified in the cp genomes of Pilea by using OGDRAW. They have a conservative quartile structure which are composed of a LSC region, a SSC region and a pair of IR regions. The genes outside the circle are transcribed in the counter clockwise direction, and the genes inside the circle are transcribed in the clockwise direction. Different colors in genes represent different functions. The dark gray area and light gray area of the inner circle represents the GC content to AT content of the genome, respectively.
Figure 2

Comparison of the Repeats in the cp genomes of 4 Pilea species. A. Types and numbers of SSRs detected in the 4 cp genomes; B. Types and numbers of the interspersed repeats in the 4 cp genomes.
Figure 3

Comparison of the borders among LSC, SSC, and IR regions of ten analyzed species. The genes around the borders are shown above or below the mainline. The JLB, JSB, JSA, and JLA represent junction sites of LSC/IRb, IRb/SSC, SSC/IRa, and IRa/LSC, respectively.
Figure 4

Comparison of the cp genomes in the 4 Pilea species by using mVISTA. The genes were represented as grey arrows on the top of the alignment. Different regions are labeled with different colors. The pink regions are "Conserved Non-Coding Sequences" (CNS), the dark blue regions are exons, and the light-blue regions are tRNA or rRNA. 50% and 100% refer to the similarity among sequences. Gray arrows above the aligned sequences represent genes and their orientation.
Nucleotide diversity (Pi) of cp genomes among the 4 Pilea species. Each black dot represents the nucleotide diversity per 500 bp. There were seven intergenic regions (petN-psbM, 0.06067; psbZ-trnG-GCC, 0.07067; trnT-UGU-trnL-UAA, 0.06433; accD-psbI, 0.06003; ndhF-rpl32, 0.06100; rpl32-trnL-UAG, 0.06800; ndhA-intron, 0.06533) and one protein-coding region (ycf1, 0.07367-0.17067) had Pi values greater than 0.06.
Figure 6

Sequence polymorphism among 79 shared cp genes of 4 Pilea species. A. Percentages of variable sites in 79 shared protein-coding genes. We used MEGA 6.0 to calculate the percentages of variable sites. Label the three genes with the highest mutation rates with *, they are ycf1 (16.62%), matK (10.54%) and ccsA (8.74%). B. The estimations of nonsynonymous (dN), synonymous (dS) substitution rates and dN/dS of 79 shared protein-coding genes. Label the four genes with the highest dN/dS values with *.
Figure 7

Phylogenetic relationships of species from Urticaceae inferred using Maximum likelihood (ML) method. The phylogenetic tree constructed using the complete cp genome sequences among the 25 cp genomes. The number at the bottom of the scale, 0.01, means that the length of the branch represents the replacement frequency of bases at each site of the genome at 0.01. Bootstrap values were calculated from 1000 replicates. Two nodes, which have BS < 100 for ML were marked specifically. The orange stars indicates the cluster of Pilea. Two taxa from Moraceae, namely, M. indica and F. carica were used as outgroups.

Supplementary Files

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

- TableS8.docx
- TableS7.docx
- TableS6.docx
- TableS5.docx
- TableS4.xlsx
- TableS3.docx
- TableS2.docx
- TableS1.docx