Chloroplast phylogenomic analysis provides insights into the evolution of the largest eukaryotic genome holder, *Paris japonica* (Melanthiaceae)

Lifang Yang¹,⁵ †, Zhenyan Yang¹ †, Changkun Liu¹, Zhengshan He³, Zhirong Zhang³, Jing Yang³, Haiyang Liu⁴, Junbo Yang³* and Yunheng Ji¹,²*

**Abstract**

**Background:** Robust phylogenies for species with giant genomes and closely related taxa can build evolutionary frameworks for investigating the origin and evolution of these genomic gigantisms. *Paris japonica* (Melanthiaceae) has the largest genome that has been confirmed in eukaryotes to date; however, its phylogenetic position remains unresolved. As a result, the evolutionary history of the genomic gigantisms in *P. japonica* remains poorly understood.

**Results:** We used next-generation sequencing to generate complete plastomes of *P. japonica*, *P. verticillata*, *Trillium govanianum*, *Ypsilandra thibetica* and *Y. yunnanensis*. Together with published plastomes, the infra-familial relationships in Melanthiaceae and infra-generic phylogeny in *Paris* were investigated, and their divergence times were calculated. The results indicated that the expansion of the ancestral genome of extant *Paris* and *Trillium* occurred approximately from 59.16 Mya to 38.21 Mya. The sister relationship between *P. japonica* and the section *Euthyra* was recovered, and they diverged around the transition of the Oligocene/Miocene (20 Mya), when the Japan Islands were separated from the continent of Asia.

**Conclusions:** The genome size expansion in the most recent common ancestor for *Paris* and *Trillium* was most possibly a gradual process that lasted for approximately 20 million years. The divergence of *P. japonica* (section *Kinugasa*) and other taxa with thick rhizome may have been triggered by the isolation of the Japan Islands from the continent of Asia. This long-term separation, since the Oligocene/Miocene boundary, would have played an important role in the formation and evolution of the genomic gigantism in *P. japonica*. Moreover, our results support the taxonomic treatment of *Paris* as a genus rather than dividing it into three genera, but do not support the recognition of *T. govanianum* as the separate genus *Trillidium*.

**Keywords:** Plastome, Phylogenomics, Giant genome, Evolution, *Paris japonica*, *Paris*, *Trillium govanianum*
Background
Angiosperms exhibit extreme diversity in genome size that is defined as the haploid nuclear DNA amount, varying by approximately 2400-fold between the smallest and largest genomes [1–4]. Although the distribution of genome size in angiosperms is strongly skewed towards small genomes (with a mean value of 1C = 5.7 Gb and a modal value of 1C = 0.6 Gb) [4], to date, five species with the genome size 1C > 100 Gb have been documented. These plant species belong to the monocotyledonous family Melanthiaceae (one species in Paris, two species in Trillium), Liliaceae (one species in Fritillaria), and eudicot family Viscaceae (one species in Viscum) [5–9], suggesting that genomic gigantism may have originated and evolved independently in only a few lineages [1, 10].

Because of the technical challenges in sequencing very large or very small genomes, insights into the mechanisms that drive the formation of genomic gigantism remain limited [9]. High-resolution and well-supported phylogenetic relationships between species with giant genomes and their closely related taxa can build evolutionary frameworks to elucidate the evolutionary history of these genomic gigantisms [9–12]. Unfortunately, a robust phylogeny for the genera Paris, Trillium and Viscum, which include genomic gigantisms, remains elusive [13–15], which impedes our understanding of the mechanisms underlying the formation and evolution of giant genomes.

Although the genome size of Polychaos dubia, a unicellular eukaryote, has been estimated to be over 670 Gb [16], this measurement is considered unreliable and inaccurate [4]. To date, the confirmed largest genome in eukaryotes has been observed in Paris japonica (Franch. et Sav.) Franch. (also known as Kinugasa japonica (Franch. et Sav.) Tatew.et Sutô.), with the 1C value of 148.88 Gb [1, 17]. This plant is a perennial herb belonging to the monocotyledonous family Melanthiaceae tribe Parideae [18, 19], and occurs natively in central and northern Honshu, Japan [20, 21]. Cytological studies revealed that P. japonica is an octoploid with a chromosome number of 2n = 8x = 40 [1, 22, 23]. Because of its distinctive characters, such as showy and white sepal, and octoploid chromosome count, P. japonica has been historically placed either in the genus Paris (section Kinugasa) [21, 23] or treated as a monotypic genus Kinugasa [20, 24]. Moreover, the evolutionary relationships of P. japonica with related taxa have remained controversial in recent analyses based on single or multiple-locus DNA sequences. An analysis using the plastid rbcL region indicated that P. japonica is a sister to the genus Trillium [25]. A combination analyses of the plastid rbcL and matK and nuclear ITS DNA regions revealed that P. japonica is closely related to the genus Daiswa (=Paris section Euthyra) [13, 26]. By contrast, two independent studies that based on the plastid psbA-trnH, trnL-F and nuclear ITS sequence data [27], and the combination of five plastid regions (atpB, rbcL, matK, ndhF and trnL-F) [28], resolved P. japonica as the sister group of the section Paris. These conflicts suggest that the relationships between P. japonica and allied taxa require further investigation.

Phylogenetic analysis using too few DNA sequences may result in a conflict between different sequence regions [29, 30]; in such a case, it is not possible to reconstruct a robust and reliable phylogeny, in particular, at low taxonomic levels [31]. Because of its high level of intra- and infra-specific sequence variation, complete plastome DNA sequences can offer valuable information for the analysis of complex evolutionary relationships in plants [32–34]. With the advent of next-generation DNA sequencing technologies, plastomes have been widely used in recent years to reconstruct robust phylogenies for several phylogenetically difficult plant taxa [31, 35–37]; these cases suggest that whole plastome sequencing may provide novel evidence to elucidate the relationships between P. japonica and allied taxa. Despite the fact that the analysis of maternally inherited DNA loci may not demonstrate the complete history of the species, the complete plastome-based phylogeny can give us some valuable information to elucidate the maternal origin and evolution of the genomic gigantisms in P. japonica.

In the current study, we used low-coverage genome shotgun sequencing [38] to generate plastomes of P. japonica, P. verticillata, Trillium govanianum, Ypsilon thibetica and Y. yunnanensis and then inferred the molecular evolution by comparing the structure and gene content to those of other published plastomes in Melanthiaceae. Then, we reconstructed the evolutionary relationships within the family to investigate the phylogenetic position of P. japonica. Finally, we dated the divergence of P. japonica to provide insights into the evolutionary history of the largest eukaryotic genome holder.

Results
Plastid genome features
The plastome of P. japonica, P. verticillata, T. govanianum, Y. thibetica and Y. yunnanensis were completely assembled. The sequencing coverage for each plastome ranged from 283× to 1086× (Additional file 2: Table S2). The gene content (Additional file 3: Table S3, Additional file 4: Table S4, Additional file 5: Table S5, Additional file 6: S6, Additional file 7: S7) and arrangement (Additional file 8: Figure S1, Additional file 9: Figure S2, Additional file 10: Figure S3, Additional file 11: Figure S4, Additional file 12: Figure S5) across the five plastomes were almost identical. The size of these newly generated
plastomes ranged from 155,957 to 158,806 bp, which exhibited a typical quadripartite structure with a pair of IRs (26,805–27,602 bp) separated by the LSC (83,635–85,301 bp) and SSC (18,337–19,586 bp) regions (Table 1). Except for the trnD-GUC that has been deleted from the plastome of _Y. thibetica_, the other plastomes encoded 114 unique genes, including 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes (Table 2).

Although the gene content and arrangement were almost identical, pseudogenization and gene loss were found to have occasionally occurred within the family Melanthiaceae. Because of the presence of several internal stop codons in coding regions, _cemA_ was identified as a pseudogene in all _Paris_ and _Trillium_ plastomes (Fig. 1a). In addition, the loss of the first exon of _rps16_ gene was found in the plastomes of _Veratrum patulum_ and _Chionographis japonica_ (Fig. 1a). Expansion of the IR regions into the _ycf1_ gene at the IR/SSC boundary occurred identically in all plastomes in Melanthiaceae, whereas their IR/LSC junctions were significantly variable. Three types of IR/LSC boundaries were observed in Melanthiaceae and outgroup taxa (Fig. 1b). The expansion of IR into the _trnH-rps19_ intergenic spacer (type III) was only found in _V. patulum_, whereas the expansion of IR into _rps19_ (type II) occurred in _Trillium cuneatum_, _T. maculatum_, and _Paris polyphylla var. chinesis_, as well as in outgroup taxa. Comparatively, characterized by the IR/LSC boundary falling into _rps3_, type I was observed in the remaining taxa (Fig. 1a).

The length of the intergenic region between _rpl23_ and _ycf2_ exhibited substantial variation among plastomes in the family Melanthiaceae, within which single-copy, duplicates and triplicates of _trnl-CAU_ were observed (Fig. 1c). Triplication of _trnl-CAU_ (type C) was observed in _P. quadrifolia_ and _P. verticillata_ (section _Paris_), whereas duplication of _trnl-CAU_ (type B) was found in _T. maculatum_. A single-copy of _trnl-CAU_ (type A) was identified in the other plastomes (Fig. 1a).

**Phylogenomic analysis and divergence estimation**

The tree topologies from both ML and BI analyses were identical. The phylogenetic relationships among the plastomes are presented in Fig. 1a. Five well-supported clades (BS = 100%, PP =1), corresponding to the five tribes (Melanthiaceae, Chionographideae, Heloniaceae, Xerophylleae, and Parideae) recognized by Zomlefer [18], were recovered. The tribe Melanthiaceae was sister to the rest of Melanthiaceae (BS = 100%, PP =1). The sister relationships between Chionographideae and Heloniaceae, as well as between Xerophylleae and Parideae, were fully supported (BS = 100%, PP =1). The intra-tribe relationships from our phylogenomic analysis are congruent with those of previous studies based on the nuclear ribosomal ITS and plastid _trnl-trnF_ regions [18]; the combination of plastid DNA sequences [28, 39]; and the plastid genome sequencing [15].

Within the tribe Parideae, the sister relationship between _Trillium_ and _Paris_ was recovered (BS = 100%, PP =1). The _Paris_ species were further grouped into three fully supported lineages (BS = 100%, PP =1) that correspond to either the three narrowly-defined genera (_Paris_ s.s., _Kinugas_ and _Daiswa_, respectively) by Takhtajan [24] or the three sections (section _Paris_, section _Kinugas_ and section _Euthyra_, respectively) circumscribed by Hara [23]. Among them, _P. japonica_ (section _Kinugas_) was sister to the section _Euthyra_ (BS = 100%, PP =1), and the section _Paris_ was sister to the clade consisting of section _Kinugas_ and section _Euthyra_. The intersectional relationships obtained here are consistent with those of a previous study [40].

Three calibration points in Melanthiaceae (Fig. 1a) suggested by previous study [41] were used to constrain the plastome-based phylogenetic tree. The results suggested that the most recent common ancestor (MRCA) for the tribe Parideae dated at approximately 59.16 Mya (95% HPD: 53.01–69.11 Mya) and the genera _Paris_ and _Trillium_ diverged from each other approximately 38.21 Mya (95% HPD: 32.17–46.84 Mya). Within the genus _Paris_, the MRCA of the section _Paris_ dated at approximately 33.71 Mya (95% HPD: 31.82–40.15 Mya), and the divergence between the monotypic section _Kinugas_ (P. _japonica_) and the section _Euthyra_ occurred approximately 20.30 Mya (95% HPD: 22.6–9.96 Mya).

**Discussion**

Robust phylogenies for species with giant genomes and allied taxa can build evolutionary frameworks to elucidate the origin and evolution of these genomic gigan
tisms [9–12]. Previous studies [13, 25–27] revealed that it is difficult to reconstruct high-resolution and well-

| Species              | Whole plastome size | LSC size | SSC size | IR size |
|----------------------|---------------------|----------|----------|---------|
| Paris japonica       | 155,957 bp          | 83,635 bp| 18,712 bp| 26,805 bp|
| P. verticillata      | 157,946 bp          | 83,710 bp| 19,586 bp| 27,325 bp|
| Trillium govanianum  | 157,379 bp          | 83,802 bp| 18,651 bp| 27,463 bp|
| Ypsilandra yunnanensis | 158,806 bp      | 85,301 bp| 18,383 bp| 27,561 bp|
| Y. thibetica         | 157,613 bp          | 84,072 bp| 18,337 bp| 27,602 bp|
Table 2: Summary of gene content in the five newly sequenced plastomes

| Species                | No. of protein-coding genes | No. of tRNA | No. of rRNA | Total |
|------------------------|----------------------------|-------------|-------------|-------|
| Paris japonica         | 80                         | 30          | 4           | 114   |
| P. verticillata        | 80                         | 30          | 4           | 114   |
| Trillium govanianum    | 80                         | 30          | 4           | 114   |
| Ypsilandra yunnanensis| 80                         | 30          | 4           | 114   |
| Y. thibetica           | 80                         | 29          | 4           | 113   |

Fig. 1 Phenylogeny, molecular dating, comparison of IR expansions and trnL-CAU copy number based on complete plastome DNA sequences from Melanthiaceae. **A** Phylogenetic relationships within Melanthiaceae based on complete cp genome sequences. Patterns of IR expansion and copy number of trnL-CAU for each species were mapped along the tree. Numbers above/under the tree branches represent BS and PP values (mean divergent ages). Arrows indicated the calibrating points for molecular dating. Horizontal blue bars on each node indicate the 95% confidence interval of divergence time. Numbers on the Time Axis indicate million year ago (Mya). **B** Three types of IR/LSC expansions were detected in the plastomes within Melanthiaceae. **C** Single-copies, duplicates, and triplicates of the trnL-CAU gene were found in the plastomes within Melanthiaceae.
supported phylogenetic relationships between *P. japonica*, the largest eukaryotic genome holder, and its allied taxa based on too few DNA sequence regions. In this study, we sequenced the whole plastomes of *P. japonica*, as well as *P. verticillata*, *Trillium govanianum*, *Ypsilandra thibetica* and *Y. yunnanensis*. Coupled with publicly available plastomes in Melanthiaceae, we performed comparative and phylogenetic analyses of whole plastomes to clarify the evolutionary relationships of *P. japonica* with its closely related taxa. This study gives us some new information about the origin and evolution of the genomic gigantisms in *P. japonica*.

**Plastome comparison**

The loss of the first exon of *rps16* was observed in the phylogenetically distinctive tribes Melanthiaceae and Chionographideae. Furthermore, the loss of *trnD-GUC* was only found in *Y. thibetica*. These results support the deduction that the loss of certain plastid genes may have independently occurred over the evolutionary history of angiosperms [32, 42]. Therefore, the loss of certain plastid genes may not provide relevant evolutionary information. However, neither gene loss nor gene relocation were observed in any of the Melanthiaceae plastomes, implying the gene content and plastome structure in the family are highly conserved.

Previous studies have revealed that the protein-coding gene *cemA* has been lost in several non-photosynthetic parasitic plants [43–45]. To our knowledge, pseudogenization of this gene in photosynthetic autotrophic angiosperms has been only detected in the closely related genera *Paris* and *Trillium* (Fig. 1a). Although its function remains unclear [46], this mutation may provide a molecular synapomorphy to recognize the tribe Parideae [47]. In addition, as proposed in a previous study [15], the lineage-specific triplcation of *trnL-CALU* in *P. quadrifolia* and *P. verticillata* could be used as a molecular synapomorphy to circumscribe the section *Paris* (Fig. 1a).

The IR/LSC boundaries of monocot plastomes generally expand into the *trnH-rps19* gene cluster and the IR expansion duplicate *trnH* gene, which differs from those of non-monocot angiosperms [47]. In this study, we identified three types of IR/LSC expansions within Melanthiaceae; of those, type II and III exhibited the typical monocot IR/LSC junctions, whereas the IR/LSC junctions of type I fell in *rps3*. Although IR/LSC expansions into the *rps19-rpl22* intergenic spacer or *rpl22* have been observed in some monocot orders, such as Asparagales, Commelinales, Ziniberales and Poales [48–50], the more progressive expansion of IR/LSC into *rps3* has only been found in Melanthiaceae to date. The phylogenetic distribution of the three types of IR/LSC boundary in the tree topology suggests that the type III can be the ancestral state in Melanthiaceae, by compared with the expansion of IR regions into *rps3* occurring in the derived tribes such as Chionographideae, Heloniadeae, Xerophyllae, and Parideae (Fig. 1a). Furthermore, the observation of type II of IR/LSC junction in *T. cuneatum*, *T. maculatum*, and *Paris polyphylla* var. *chinensis* may have been resulted from a secondary slippage of IR regions from *rps3* to *rps19*.

**Phylogeny inferences**

Our phylogenomic analysis recovered five well-supported lineages (BS = 100%, PP = 1) within Melanthiaceae, which correspond to the five tribes recognized by Zomlefer [18]. The evolutionary relationships recovered in this study are consistent with those of previous investigations [18, 28, 40, 51] but with higher branch support (BS = 100%, PP = 1). The results further justify that whole plastid genome sequencing can improve the phylogenetic resolution in a certain lineage [33, 34].

Our expanded sampling of the plastomes in Parideae provided an opportunity to reconstruct a robust intra-generic phylogeny in the tribe. The basal divergence in Parideae occurred approximately 38.21 Mya, forming two fully supported lineages (*Paris* and *Trillium*) in the tree topology (BS = 100%, PP = 1). The two genera share synapomorphies, including a single whorl of net-veined leaves presenting at a stem apex, a stem apex bearing a solitary flower, and a chromosome base number *n* = 5 [16]. Within the clade *Paris*, the three sections (section *Paris*, section *Kinugasa*, and section *Euthyra*) outlined by Hara [23] as well as the three narrowly defined genera *Paris* s.s., *Daiwa* and *Kinugasa* by Takhtajan [24] were each recovered as monophyletic clades with strong support (BS = 100%, PP = 1) in both the ML and BI analyses. Given that species in the *Paris* clade share the morphological synapomorphies of flowers and leaves, 4- to 15-merous compared with the trimerous condition of *Trillium* [27], we correspondingly prefer to accept the taxonomic treatment of *Paris* as a single genus [21, 23] rather than in three separated genera [24].

Since a previous study had not included the plastome of *P. japonica* in its phylogenetic analysis, its evolutionary relationships with other *Paris* species remained unresolved [15]. Both ML and BI analysis identically indicated that *P. japonica* (section *Kinugasa*) is a sister to the section *Euthyra*, which is congruent with the analyses of the plastid *rbcL*, *matK* and *trnl-trnf* regions [13, 26, 40]. However, the relationships recovered by our data largely differ from the results of combination analysis of plastid *psbA-trnH* and *trnl-F* and nuclear ITS sequences [27], and plastid *atpB*, *rbcL*, *matK*, *ndhF* and *trnl-F* regions [28]. It is noteworthy
that, the well-supported sister relationship between *P. japonica* and the section *Euthyra* (BS = 100%, PP = 1) recovered in this study, can be also justified by the morphological synapomorphies that they share, such as a thick rhizome and angular ovary, in contrast to the long and slender rhizome and rounded ovary species of the section *Paris* (Fig. 2). In addition, the unusual morphological characteristics of the species (i.e., the showy, white sepals, and octoploid chromosome number) justify the taxonomic treatment of *P. japonica* as a distinctive section within the genus *Paris* by Hara [23].

Our data not only recovered the evolutionary backbone in *Paris* but also offered evidence to clarify disputes about the phylogenetic position of *T. govanianum*, which occurs natively in the Himalayan mountains. Although *T. govanianum* has a trimerous flower and leaves like those of *Trillium* species, it shares morphological features, such as narrow sepals and filiform petals, with *Paris* species (Fig. 3). Accordingly, *T. govanianum* was recognized as a separate genus *Trillidium* [13, 52]. However, neither the ML nor BI tree topology separated *T. govanianum* from the *Trillium* species but grouped them into a well-supported clade (BS = 100%, PP = 1). It is notable that similar finding has been shown in the phylogenetic analysis based on five plastid DNA regions that has a more extensive taxon sampling of Melanthiaceae [28]. Taken together, the results suggest that *T. govanianum* should remain in the genus *Trillium* and deny the recognition of the genus *Trillidium*.

**Insights into the origin and evolution of the genomic gigantism in *Paris japonica***

The robust phylogeny reconstructed in the current study provided insights into the origin and evolution of the genomic gigantism in *P. japonica*. Most species in Melanthiaceae possess small or very small genomes, while large or giant genomes have been exclusively found in the two genera: *Paris* and *Trillium* [40]. Character reconstruction revealed that a genome size increase (more than four-fold) possibly occurred after the divergence of *Xerophylleae* and *Parideae*, but before the differentiation between *Paris* and *Trillium* [40]. Molecular dating indicated that the stem age and crown age of *Parideae* were approximately 59.16 Mya and 38.21 Mya, respectively, suggesting that the massive genome expansion would have lasted for a long period of approximately 20 million years. During this period, the ancestral genome of extant *Paris* and *Trillium* would have gradually expanded, implying that the genome size increase in *Parideae* could be the slow accumulation over tens of millions of years as a previous study proposed [40].

The phylogenomic analyses indicated that the section *Paris* is sister to the clade including *P. japonica* (section *Kinugasa*) and the section *Euthyra*. These relationships suggest that the formation of a giant genome in *P. japonica* most likely took place after the divergence of the sections *Euthyra* and *P. japonica*. Except for *P. japonica*, two species (*T. × hagae* and *T. rhombilolium*) with genome sizes 1C > 100 Gb have been found in the genus *Trillium* [5, 6, 9]. As we did not obtain samples of these two plants, their phylogenetic positions within *Trillium* remain unclear. Nevertheless, the evolutionary relationships of *P. japonica* with related taxa recovered in the study reveal that the formation of the giant genomes in *P. japonica* and *Trillium* species may have been independent events.

The coalescence of the plastomes of *P. japonica* and the section *Euthyra* occurred around the transition of the Oligocene/Miocene (20.30 Mya, 95% HPD: 34.64–9.96 Mya), when the opening of the Japan Sea separated the Japan Islands from the continent of Asia [53]. Although *P. japonica* and the section *Euthyra* are closely related, they
occupy distinct distributions: *P. japonica* is endemic to Japan, whereas species from the section *Euthyra* are chiefly distributed in subtropical China and the Himalayas [23]. Hence, the divergence of *P. japonica* and the section *Euthyra* may have been triggered by the isolation of the Japan Islands from the continent of Asia.

Notably, the genome size of *P. japonica* is approximately 2–3 folds larger than that of those species belonging to the section *Euthyra* [40]. A line of evidence justifies that the genome size variation in plants is under selective constrains and has not evolved by a pure drift process [54–56]. As a result, genome size can be strictly related to the environment and ecology of a species [57]. In general, plants with larger genomes share some morphological traits, such as large body and stomata size [58]. Due to the drought susceptibility of the plants with large stomata, only those species occurring in humid habitats can sustain larger genomes [56, 58]. Compared with the monsoonal climate which is characterized by obvious precipitation seasonality in subtropical China and the Himalayas [59, 60], the maritime climate of the Japan islands [61] would create relatively more humid habitats that facilitate the evolution of *P. japonica* toward genomic gigantism.

**Conclusions**

The evolutionary relationships of the largest eukaryotic genome holder, *P. japonica*, with its closely related taxa were investigated by comparative and phylogenetic analyses of their complete plastome DNA sequences. Comparative analysis across plastomes in Melanthiaceae revealed that their structures and gene contents are highly conserved and provided molecular synapomorphies for some lineages of Parideae. Phylogenomic analysis and molecular dating recovered the evolutionary backbone of *Paris* and thus elucidated the phylogenetic position of *P. japonica*. The tree topologies and molecular dating indicated that the expansion of the ancestral genome of extant *Paris* and *Trillium* was probably a gradual process lasting for approximately 20 million years; the divergence of *P. japonica* and the section *Euthyra* may have been triggered by the opening of the Japan Sea, which separated the Japan Islands from the continent of Asia around the transition of the Oligocene/Miocene (20.30 Mya). This long-term separation would have played an important role in the formation and evolution of genomic gigantism in *P. japonica*. The phylogenetic position of *P. japonica* implies that the giant genomes of *Paris* and *Trillium* may have formed and evolved independently, even though the two genera
are closely related. In addition, our phylogenomic analysis strongly supports the taxonomic treatment of *Paris* as a genus rather than dividing it into three genera, but did not support the recognition of *T. govanianum* as the separate genus *Trillidium*.

**Methods**

**Plant material and shotgun sequencing**

Leaf tissues of *P. japonica*, *P. verticillata*, *T. govanianum*, *Y. thibetica* and *Y. yunnanensis* were collected in the field and then dried with silica gel (one individual per species). The vouchers were identified by Dr. Yunheng Ji and deposited at the herbarium of Kunming Institute of Botany, Chinese Academy of Sciences (KUN); the voucher information is presented in Table 3. Genomic DNA was extracted from ~20 mg of leaf tissue using a modified CTAB method [62]. Approximately 5 μg of purified genomic DNA was sheared by sonication. Paired-end libraries with an average insert size 350 bp were prepared using a TruSeq DNA Sample Prep Kit (Illumina, Inc., USA) according to the manufacturer’s protocol. Shotgun sequencing was performed on the Illumina HiSeq 2000 platform.

**Plastome assembly, annotation and comparison**

Raw Illumina reads were filtered by NGS QC tool kit [63] to remove adaptors and low-quality reads. The pipeline developed by Jin et al. [64] was used for de novo plastome assembly. The clean reads of *Paris* species, *T. govanianum* and *Ypsilandra* species were mapped onto the reference plastomes of *P. quadrifolia* (Genbank accession: KX784051), *T. tschonoskii* (Genbank accession: KR780076) and *Heloniopsis tubiflora* (Genbank accession: KM078036) using the Bowtie v2.2.6 software [65] with its default parameters and preset options. All of the plastid-like reads were assembled into contigs by SPAdes v3.10.1 with the default parameters and preset options. The results of de novo assembly were visualized and edited with Bandage v.8.0 [67]. The verifying plastomes were annotated by Dual Organellar Genome DRAW program [68]. The plastomes were checked manually. All of the identified tRNA was verified by tRNAscan-SE v1.21 [70], with the preset parameters. Functional classification of the plastid genes was determined by referring to the online database CpBase (http://rocaplab.ocean.washington.edu/old_website/tools/cpbase). The maps of plastomes were constructed with the Organellar Genome DRAW program [71].

The general features of plastome, such as structural rearrangements, gene loss/pseudogenization, gene duplication, and expansion/contraction of the IR regions, have provided evolutionary information in previous studies [15, 32, 72]. Therefore, we performed comparisons of these features among Melanthiaceae plastomes. The gene content and arrangement were visualized and compared with the MUMmer 3.0 program [73]. The boundaries of the LSC, IR, and SSC regions in each plastome were compared using Geneious v10.2.3 [69].

**Phylogenomic analysis**

To examine the phylogenetic position of *P. japonica*, 24 plastomes representing wide phylogenetic diversity in the family Melanthiaceae were included in the phylogenomic analysis (Additional file 1: Table S1). The plastomes of *Campynema lineare*, *Fritillaria cirrhosa*, *Luzuriaga radicans* and *Smilax china* were used to root the tree. Of those, five plastomes were newly generated in the current study (Table 3), and the rest plastomes were downloaded from the NCBI database (Additional file 1: Table S). The complete plastome DNA sequences were aligned using MAFFT [74] integrated in Geneious v.10.2.3 [69], with manual adjustment if necessary. The phylogenomic analyses were carried out with the standard Maximum Likelihood (ML) and Bayesian Inference (BI) methods. ML analyses were performed using RAxML-HPC BlackBox v8.1.24 [75] with 1000 replicates of rapid bootstrapping (BS) under the GTR-GAMMA model. The search of the best-scoring ML tree and rapid bootstrapping were performed in a single run. The choice of the best nucleotide sequence substitution model for BI analysis was determined using Modeltest v3.7 [51] with the Akaike Information Criterion [76]. BI was performed with MRBAYES v.3.1.2 [77] using the model (TVM + I + F).

**Table 3:** Plastomes newly generated in this study with taxon, source, voucher information, and GenBank accessions

| Species                | Source of plant material | Voucher (Herbarium) | Genbank accession |
|------------------------|--------------------------|---------------------|-------------------|
| *Paris japonica*       | Chubu, Honshu, Japan     | J. Maruta s. n. (KUN) | MF796668          |
| *P. verticillata*      | Jamusi, Heilongjiang, China | L. X. Wang s. n. (KUN) | MF796669          |
| *Trillium govanianum*  | Dingri, Tibet, China     | S. K. Chen 1,289,634 (KUN) | MF796670          |
| *Ypsilandra yunnanensis* | Gongshan, Yunnan, China | Y. Ji 2,007,014 (KUN) | MF796672          |
| *Y. thibetica*         | Nanchuan, Chongqing, China | Y. Ji 2,013,031 (KUN) | MF786671          |
Two independent parallel Markov Chain Monte Carlo (MCMC) runs with tree sampling every 100 generations for one million generations, with the first 25% discarded as burn-in, were conducted. Stationarity was considered to be reached when the average standard deviation of the split frequencies was < 0.01. The posterior probability values (PP) were determined from the remaining 0.75 million trees.

Molecular dating

To date, no fossils have been identified for the family Melanthiaceae and its close relatives. Calibrated by 17 fossils across the monocots and major clades of angiosperms, a previous study [41] revealed that the crown age of family Melanthiaceae was approximately 84.8 Mya, while the clades Parideae-Xerophyllidaceae and Chionographidaceae-Heloniadae diverged approximately 74 Mya, and the tribes Parideae and Xerophyllidaceae split approximately 52.3 Mya. We used these events to calibrate the phylogenetic tree (Fig. 1a).

Molecular dating was performed using the MCMCTREE v4.9c program integrated in the PAML program package [78]. The ML tree topology was used to estimate the divergence times of nodes. The independent-rates molecular clock was chosen as the clock model, and HKY85 was selected as the substitution model. The root age was set as less than 100 Mya. The divergence of Melanthiaceae was calibrated with a minimum age of 84.8 Mya. The node uniting Parideae-Xerophyllidaceae and Chionographidaceae-Heloniadae was set to a minimum age of 74 Mya, while the divergence of Parideae and Xerophyllidaceae was set to a minimum age of 52.3 Mya. Other parameters were defined as their defaults. MCMC chains were run for 10,100,000 iterations. The first 100,000 iterations were discarded as burn-in, and trees were sampled every 10 iterations until 1,000,000 samples were gathered.

Additional files

- Additional file 1: Table S1. Plastomes included in the phylogenetic analyses with GenBank accession. (DOCX 19 kb)
- Additional file 2: Table S2. Summary of the Illumina sequencing results of Paris japonica, P. verticillata, Trillium govanianum, Ypsilandra thibetica and Y. yunnanensis. (DOCX 14 kb)
- Additional file 3: Table S3. List of the genes identified in the plastome of Paris japonica. (DOCX 16 kb)
- Additional file 4: Table S4. List of the genes identified in the plastome of Paris verticillata. (DOCX 16 kb)
- Additional file 5: Table S5. List of the genes identified in the plastome of Trillium govanianum. (DOCX 16 kb)
- Additional file 6: Table S6. List of the genes identified in the plastome of Ypsilandra thibetica. (DOCX 16 kb)

Acknowledgements

We thank the three anonymous reviewers for their critical comments in improving the manuscript. We are also grateful to Prof. Liangxin Wang, and Dr. J. Maruta for providing some samples used in this study.

Authors’ contributions

YJ and Y2 conceived and designed the research, LY, ZY, CL, ZZ, and Y1 collected and analyzed the data. YJ, LY and ZY prepared the manuscript. JY1 and HL discussed the results and revised the manuscript. All of the authors read and approved the manuscript.

Funding

The authors would like to thank the financial support from the Major Program of National Natural Science Foundation of China (31590823), the National Natural Science Foundation of China (31872673), the NSFC-Joint Foundation of Yunnan Province (U1802287), the Large-scale Scientific Facilities of the Chinese Academy of Sciences (No. 2017-LSF-GBOWS-02), and the Guiding Program of Interdisciplinary Studies (Grant No. KIB2017004) from the Kunming Institute of Botany (CAS) in design of the study, and collection, analysis and interpretation of data, as well as in writing the manuscript.

Availability of data and materials

The complete cp genome sequences of P. japonica, P. verticillata, Trillium govanianum, Ypsilandra thibetica and Y. yunnanensis are available at GenBank under the accession numbers MF796668–MF796672. The data used in the analysis are included within the article and the additional files.

Ethics approval and consent to participate

Collection of all samples completely complies with national and local legislation permission. This study did not involve any endangered or protected species, and the plant samples used in the study were not collected from national park or natural reserve. According to national and local legislation, no specific permission was required for collecting these plants.

Consent for publication

Not applicable.
Competing interests
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author details
1 Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, People’s Republic of China. 2 Yunnan Key Laboratory for Integrative Conservation of Plant Species with Extremely Small Population, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China. 3 Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China. 4 State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China. 5 School of Life Science, Yunnan University, Kunming, China.

Received: 26 October 2018 Accepted: 10 June 2019
Published online: 04 July 2019

References
1. Pellizzer J, Fay MF, Leitch IJ. The largest eukaryotic genome of them all? Bot J Linn Soc. 2010;164:10–5.
2. Pellizzer J, Kelly LJ, Magdalena C, Leitch IJ. Insight into the dynamics of genome size and chromosome evolution in the early diverging angiosperm lineage Nymphalaeas (water lilies). Genome. 2013;56:437–49.
3. Fedorov NV. Transposable elements, epigenetics, and genome evolution. Science. 2012;338:758–66.
4. Dodsword S, Leitch AR. Genome size diversity in angiosperms and its influence on gene space. Curr Opin Genet Dev. 2015;35:73–8. https://doi.org/10.1016/j.gde.2015.10.006.
5. Leitch IJ, Chase MW, Leitch AR. Genome size diversity and evolution in land plants. In: Li H, editor. The phylogeny of the genus Paris L. In: Li H, editor. The phylogeny of the genus Paris L. Beijing: Science Press; 1998. p. 8–65.
6. Haga T. Chromosome complement of Kinugasa japonica with special reference to its origin and behavior. Cytologia. 1938;8:137–41. https://doi.org/10.1508/cytologia.8.137.
7. Yang H. Variations in Paris polyphylla smith, with reference to other Asiatic species. J Fac Sci Univ Tokyo Sect III. 1969;10:141–80.
8. Takhtajan A. A revision of Daiswa (Trilliaceae). Brittonia. 1983;35:255–70. https://doi.org/10.2307/2806025.
9. Kato H, Terauchi R, Utech FH, Kawano S. Molecular systematics of the Trilliaceae sensu lato as inferred from rbcL sequence data. Mol Phylogenet Evol. 1995;4:184–93. https://doi.org/10.1006/mpev.1995.1018.
10. Osaloo SK, Kawano S. Molecular systematics of Trilliaceae II. Phylogenetic analyses of Trillium and its allies using sequences of rbcL and matK genes of cpDNA and internal transcribed spacers of 18S and 25S rDNA. Plant Species Biol. 1999;14:75–94. https://doi.org/10.1046/j.1442-1849.1999.0009.x.
11. Ji Y, Fritsch PW, Li H, Xiao T, Zhou Z. Phylogeny and classification of Paris (Melanthiaceae) inferred from DNA sequence data. Ann Bot. 2006;98:245–56. https://doi.org/10.1093/aob/mcl095.
12. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 2009;7:84. https://doi.org/10.1186/1471-2105-9-74.
13. Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller R, Guisinger-Bellian M, et al. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. Proc Natl Acad Sci U S A. 2007;104:19369–74. https://doi.org/10.1073/pnas.0709121104.
14. Moore MJ, Bell CD, Solits PS, Solits DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci U S A. 2007;104:19363–6. https://doi.org/10.1073/pnas.0708721104.
15. Moore MJ, Solits PS, Bell CD, Burleigh JG, Solits DE. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc Natl Acad Sci U S A. 2010;107:4623–8. https://doi.org/10.1073/pnas.0907801107.
16. Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JD. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? Ann Bot. 2014;113:119–33. https://doi.org/10.1093/aob/mct264.
17. Stull GW, Dunac RD, Solits DE, Solits PS. Resolving basal lamioid phylogeny and the circumscription of Icacinaceae with a plastome-scale data set. Am J Bot. 2015;102:1794–813. https://doi.org/10.3732/ajb.1500097.
18. Attigala L, Wysoki WP, Derval MR, Clark LG. Phylogenetic estimation and morphological evolution of Arundinarieae (Bambusoideae: Poaceae) based on plastome phylogenomic analysis. Mol Phylogenet Evol. 2016;101:111–121. doi: https://doi.org/10.1016/j.ympev.2016.05.008.
19. Staub SC, Parks M, Weitemier K, Fishbein M, Cronn R, Liston A. Navigating the tip of the genomic iceberg: next-generation sequencing for plant systematics. Am J Bot. 2012;99:349–64. https://doi.org/10.3732/ajb.1100355.
20. Kim JS, Hong JK, Chase MW, Fay MF, Kim JH. Familial relationships of the tribe Melanthieae (Liliales, Melanthiaceae), with emphasis on Zigadenus: evidence from ITS and rnlF sequence data. Am J Bot. 2001;88:657–69. https://doi.org/10.2307/3558411.
21. Angiosperm Phylogeny Group. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181:1–20.
22. Takewaki M, Sutô T. On the new genus kinugasa. Trans Sapporo Not Hist Soc. 1935;14:34–6.
23. Li H. The phylogeny of the genus Paris L. In: Li H, editor. The genus Paris L. Beijing: Science Press; 1998. p. 8–65.
24. Haga T. Chromosome complement of Kinugasa japonica with special reference to its origin and behavior. Cytologia. 1938;8:137–41. https://doi.org/10.1508/cytologia.8.137.
25. Yang H. Variations in Paris polyphylla smith, with reference to other Asiatic species. J Fac Sci Univ Tokyo Sect III. 1969;10:141–80.
26. Takhtajan A. A revision of Daiswa (Trilliaceae). Brittonia. 1983;35:255–70. https://doi.org/10.2307/2806025.
27. Pellizzer J, Kelly LJ, Leitch IJ, Zomlefer WB, Fay MF. A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. New Phytol. 2014;201:1484–97.
