Comparison of the chloroplast genomes and phylogenomic analysis of Elaeocarpaceae

Yihui Wang[1,2,*], Yifei Xie[1,3,*], Jiayi Jin[1], Jinyue Li[4], Xiangdong Qiu[1], Yang Tong[1], Zhongyang Li[1,3], Zhixiang Zhang[5] and Wenling Lai[1,3]

[1] School of Life Sciences, Gannan Normal University, Ganzhou, P.R. China
[2] School of Landscape Architecture, Beijing Forestry University, Beijing, P.R. China
[3] Key Laboratory of Nanling Plant Resources Conservation and Utilization, Ganzhou, P.R. China
[4] Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Xishuangbanna, P.R. China
[5] School of Ecology and Nature conservation, Beijing Forestry University, Beijing, P.R. China
* These authors contributed equally to this work.

Corresponding author
Yifei Xie, xiey-f@foxmail.com

ABSTRACT

Background: Elaeocarpaceae is a vital family in tropical and subtropical forests. Compared with the important position of Elaeocarpaceae species in forest ecosystem and the concern of medicinal value, the most research on Elaeocarpaceae are classification and taxonomy. Molecular systematics has corrected the morphological misjudgment, and it belongs to Oxalidales. Phylogenetic and divergence time estimates of Elaeocarpaceae is mostly constructed by using chloroplast gene fragments. At present, although there are reports on the chloroplast structure of Elaeocarpaceae, a comprehensive analysis of the chloroplast structure of Elaeocarpaceae is lacking.

Methods: To understand the variation in chloroplast sequence size and structure in Elaeocarpaceae, the chloroplast genomes of nine species were sequenced using the Illumina HiSeq 2500 platform and further assembled and annotated with *Elaeocarpus japonicus* and *Sloanea sinensis* (family Elaeocarpaceae) as references. A phylogenomic tree was constructed based on the complete chloroplast genomes of the 11 species representing five genera of Elaeocarpaceae. Chloroplast genome characteristics were examined by using Circoletto and IRscope software.

Results: The results revealed the following: (a) The 11 sequenced chloroplast genomes ranged in size from 157,546 to 159,400 bp. (b) The chloroplast genomes of *Elaeocarpus*, *Sloanea*, *Crinodendron* and *Vallea* lacked the *rpl*32 gene in the small single-copy (SSC) region. The large single-copy (LSC) region of the chloroplast genomes lacked the *ndh*K gene in *Elaeocarpus*, *Vallea stipularis*, and *Aristotelia fruticosa*. The LSC region of the chloroplast genomes lacked the *inf*A gene in genus *Elaeocarpus* and *Crinodendron patagua*. (c) Through inverted repeat (IR) expansion and contraction analysis, a significant difference was found between the LSC/IRB and IRA/LSC boundaries among these species. *Rps*3 was detected in the neighboring regions of the LSC and IRb regions in *Elaeocarpus*. (d) Phylogenomic analysis revealed that the genus *Elaeocarpus* is closely related to *Crinodendron patagua* on an independent branch and *Aristotelia fruticosa* is closely related to *Vallea stipularis*, forming a clade with the genus *Sloanea*. Structural comparisons showed that Elaeocarpaceae diverged at 60 Mya, the genus *Elaeocarpus* diverged 53 Mya and that

the genus *Sloanea* diverged 0.44 Mya. These results provide new insight into the evolution of the Elaeocarpaceae.

INTRODUCTION

In land plants, most chloroplast genomes' genome is approximately 100–220 kbp in size and has a quadripartite structure, including one large single-copy (LSC) region, one small single-copy (SSC) region, and a pair of inverted repeat (IR) regions (*Bock, 2007*; *Song et al., 2017*). These regions are involved in photosynthesis, transcription, and translation, among other functions (*Gao, Su & Wang, 2010*). With the increase in complete chloroplast genome data, comparative analysis of chloroplast genomes has been widely applied (*Wu, 2016*). Some lineages, such as ferns (*Roper et al., 2007*; *Karol et al., 2010*), gnetophytes (*McCoy et al., 2008*; *Wu et al., 2009*), multiple angiosperm families (*Goremykin et al., 2003*; *Cai et al., 2006*), and nonphotosynthetic plants (*Wicke et al., 2016*), have lost some genes. For example, *ycf*1, *ycf*2 and *acc*D have been lost in the family Poaceae (*Guisinger et al., 2010*), and *rpl*22, *inf*A and *acc*D have been lost in legumes, Lemnoideae, and Acoraceae, respectively (*Doyle, Doyle & Palmer, 1995*; *Goremykin et al., 2005*; *Wang & Messing, 2011*). The *ndh* genes loss events were also detected in heterotrophic plants, utotrophic orchids, gnetophytes and Pinaceae, and pseudogenization occurred in heterotrophic plants (*Wakasugi et al., 1994*; *Wickett et al., 2008*; *Braukmann, Kuzmina & Stefanovic, 2009*; *Barrett et al., 2014*; *Kim et al., 2015*; *Wicke et al., 2016*). In recent years, phylogenomics has shown great advantages in plant phylogenetic research based on chloroplast genomes, providing resolutions for the phylogenies of some taxonomically difficult groups of plants.

Elaeocarpaceae Juss. is a medium-sized family of angiosperms comprising 12 genera and 615 species of trees that grow in tropical and subtropical forests (*Coode, 2004*; *Christenhusz & Byng, 2016*). Most studies suggest that Elaeocarpaceae is a sister group to Cephalotaceae and *Brunellia* based on *trn*L-*trn*F and nuclear *ITS* regions (*Crayn, Rossetto & Maynard, 2006*; *Magallon et al., 2015*; *Harris & Davies, 2016*). *Heibl & Renner (2012)* phylogenetic analysis indicated that Elaeocarpaceae is sister to Cunoniaceae relied on *rbc*L region, *trn*L-*trn*F region and *ITS* region. Molecular phylogenies provide a robust evidence that Elaeocarpaceae has a closest relationship with the clade comprising Cunoniaceae and Cephalotaceae (*Phoon, 2015*). Besides, the differentiation within Elaeocarpaceae is ambiguous. According to the original research, Elaeocarpaceae diverged in the early Cretaceous (120 Mya) based on *trn*L-*trn*F and nuclear *ITS* regions (*Crayn, Rossetto & Maynard, 2006*). Combined nuclear and chloroplast DNA sequences, *Heibl & Renner (2012)* study indicated that Elaeocarpaceae diverged in the late Cretaceous (67 Mya) based on three markers (*rbc*L region, *trn*L-*trn*F region and *ITS* region). *Phoon (2015)* study also showed that Elaeocarpaceae and its sister lineage (Cunoniaceae + Cephalotaceae) diverged in the late Cretaceous (85 Mya) based on four regions (*psb*A-*trn*H intergenic spcer, *trn*L-*trn*F region, *trn*V-*ndh*C intergenic spcer and nuclear *Xdh*). With the increasing

Table 1 Sampled species of Elaeocarpaceae and their voucher specimens.

No.	Species	Herbarium	Voucher	Geographic origin	Accession number in GenBank
1	*Aristotelia fruticosa* Hook.f.	K	781	The Royal Botanic Gardens	MT982368
2	*Crinodendron patagua* Molina	K	652	The Royal Botanic Gardens	MT982369
3	*Vallea stipularis* L.f.	K	654	The Royal Botanic Gardens	MT982370
4	*Elaeocarpus angustifolius* Blume	BJFC	140942	Guangxi Academy of Forestry	MW242787
5	*Elaeocarpus hainanensis* Oliver	GNNU	PVHJX014291	Diaoluo Mountain, Hainan	MW602804
6	*Elaeocarpus japonicus* Sieb. et Zucc.	BJFC	160730004	Wugong Mountain, Jiangxi	MT985378
7	*Elaeocarpus japonicus* var. *yunnanensis* C. Chen & Y. Tang	BJFC	XW1746	Wenshan, Yunnan	MW242788
8	*Sloanea sinensis* (Hance) Hemsl.	BJFC	XW1956	Wenshan, Yunnan	MW004670
9	*Sloanea cordifolia* K. M. Feng ex H. T. Chang	BJFC	XW1958	Wenshan, Yunnan	MW242789
10	*Sloanea dasycarpa* (Benth.) Hemsl.	BJFC	XZ581	Wenshan, Yunnan	MW242790
11	*Sloanea longiaculeatae* Y. F. Xie & Z. X. Zhang	BJFC	XW1986	Wenshan, Yunnan	MW242791

number of genomic data about Elaeocarpaceae on NCBI, it is necessary to conduct comparison of the chloroplast genomes research and divergence time within Elaeocarpaceae (*Wang, Zhang & Xie, 2021*; *Weng et al., 2021*). In addition, genetic differentiation within Elaeocarpaceae associated with microsatellite loci is common in many species such as *E. photiniifolia* and *Tetratheca ericifolia* (*McPherson et al., 2008*; *Sugai et al., 2012*, *2013*; *Anthony, Allcock & Krauss, 2016*).

Although the taxonomy of Elaeocarpaceae belonged to Oxalidales is widely acknowledged, the relationships within Oxalidales need further study (*Savolainen et al., 2000*; *Soltis et al., 2000*; *Byng et al., 2016*). Furthermore, the age of genera and relationships within Elaeocarpaceae are incongruent in previous studies based on multigene phylogenies (*Crayn, Rossetto & Maynard, 2006*; *Heibl & Renner, 2012*; *Phoon, 2015*), hence the age of the genera within Elaeocarpaceae have not been adequately examined.

Here we used the whole chloroplast genome sequences to further explore phylogenetic relationships within Elaeocarpaceae and other relative families in detail. This study aims to (a) test genetic category between different genus within Elaeocarpaceae, (b) determine the relationships of Elaeocarpaceae within Oxalidales, (c) use the molecular data together with the data from the palaeobotanical literature to infer divergence dates and the biogeographic history of the major clades within Oxalidales and Elaeocarpaceae.

MATERIALS AND METHODS

Plant material and chloroplast genome sequencing

Leaf materials were sampled from nine species representing five genera of Elaeocarpaceae and collected from field in China and the Royal Botanic Gardens (Table 1). Voucher specimens of the collection were deposited at the Museum of Gannan Normal University, Nanling Herbarium (GNNU; Director: Yifei Xie, xie.yifei2018@gmail.com), Museum of

Table 2 Species sequence numbers from the NCBI database.

Species	Accession number in GenBank	Accession number in SRA
* *Aristotelia fruticosa*	MT982368	SRR12599405
* *Crinodendron patagua*	MT982369	SRR12599428
* *Elaeocarpus angustifolius*	MW242787	SRR12998754
* *Elaeocarpus hainanensis*	MW602804	SRR13423273
* *Elaeocarpus japonicus*	MT683335	SRR12574443
* *Elaeocarpus japonicus* var. *yunnanensis*	MW242788	SRR13003726
* *Sloanea cordifolia*	MW242790	SRR13003865
* *Sloanea longiaculeatae*	MW242791	SRR13004978
* *Sloanea dasycarpa*	MW242790	SRR13002231
* *Sloanea sinensis*	MW004670	SRR12599358
* *Vallea stipularis*	MT982370	SRR12599429
Averrhoa carambola	KX364202	
Brunellia antioquensis	MN615725	
Brunellia trianae	MN585217	
Cephalotus follicularis	NC042597	
Euonymus maackii	MW771518	
Euonymus schensianus	NC036019	
Oxalis corniculata	NC051971	
Oxalis drummondii	NC043802	
Rourea microphylla	MT537171	

Note:
* Newly published species sequences.

Beijing Forestry University (BJFC) and The Royal Botanic Gardens (K). Total genomic DNA was extracted using the magnetic bead method and then sent to Sino Geno Max Company for next-generation sequencing using the Illumina HiSeq (TM) 2500 platform in Beijing, China and the read length generated from the Illumina platform was 150 bp. The raw data were filtered by cutadapt version 1.9.1 and trimmed by Trimmomatic version 0.39 to remove low-quality bases with the parameters in S1 (*Martin, 2011*; *Bolger, Lohse & Usadel, 2014*). Then we obtained clean data and uploaded them to the NCBI SRA database in fastq format (Table 2).

Genome annotation and comparison

The clean reads were used to assemble the chloroplast genome sequence by GetOrganelle version 1.7.7.0 with the parameters in S1 (*Bankevich et al., 2012*; *Jin et al., 2020*). A circular chloroplast genome was generated after filtered De Bruijn graphs were viewed and edited using Bandage (*Wick et al., 2015*; *Song et al., 2017*). We used Plastid Genome Annotator to annotate the assembled chloroplast genomes using as reference *E. japonicus* (MT985378; *Qu et al., 2019*). The annotated chloroplast genomes have been submitted to GenBank (Table 2). Schematic diagrams of all nine chloroplast genomes were drawn by the Organellar Genome DRAW tool (*Lohse et al., 2013*), and a map of shared protein-coding genes was drawn by a Venn diagram viewer (http://jvenn.toulouse.inra.fr/app/example.

html; *Philippe et al., 2014*). mVISTA online tools (https://genome.lbl.gov/vista/mvista/about.shtml) were used to determine chloroplast genome similarity among Elaeocarpaceae (*Frazer et al., 2004*). The similarity, rearrangement and inversion of gene blocks were analyzed by Circoletto (http://tools.bat.infspire.org/circoletto/; *Darzentas, 2010*). IRscope (https://irscope.shinyapps.io/irapp/) was used to assess IR expansion and contraction in the evolution of chloroplast genomes (*Ali, Jaakko & Peter, 2018*).

Phylogenomics and molecular clock dating analysis

To infer phylogenetic relationships within the Elaeocarpaceae and other related families, 20 species of six families including Elaeocarpaceae, Cephalotaceae, Brunelliaceae, Oxalidaceae and Connaraceae were compared. The genomes from the six families included 11 new chloroplast genomes and nine published complete chloroplast genomes (Table 2), as that of *Euonymus schensianus* (NC036019) and *Eu. maackii* (MW771518), which was obtained from the NCBI database and treated as the outgroup (*Baker et al., 2021*; *Li et al., 2021*). For the species tree, Bayesian inference (BI) analyses were performed on data sets of 20 chloroplast genome sequences. Using MAFFT version 7.490 to compare the whole genome matrix by '–auto' strategy and normal alignment mode (*Katoh & Standley, 2013*) and then ambiguously aligned fragments were removed using Gblocks version 0.91b with the following parameter settings: minimum number of sequences for a conserved/flank position (11/11), maximum number of contiguous non-conserved positions (eight), minimum length of a block (10), allowed half of gap positions (*Talavera & Castresana, 2007*). Bayesian inference (BI) was performed using MrBayes version 3.2.6 (*Ronquist & Huelsenbeck, 2003*). The best-fitting DNA substitution model according to the Bayesian information criterion (BIC), GTR (General Time Reversible) + F (Felsenstein) + I (proportion of Invariable sites), was identified by using jModelTest version 2.1.10 (*Darriba et al., 2012*; *Guindon, Gascuel & Rannala, 2003*). Markov chain Monte Carlo simulations (MCMC) were run for 10,000,000 generations. The BI analysis started with a random tree and sampled trees every 1,000 generations. The first 25% of the trees were discarded as burn-in, and the remaining trees were used to generate a majority-rule consensus tree. Besides, we also estimated a maximum likelihood (ML) phylogeny for the genera in RAxML v8.0.0 (*Stamatakis, Hoover & Rougemont, 2008*), on the CIPRES web server (www.phylo.org). We used the default settings, including a TVM (Transversion model) + R3 (Free Rate three) + F (Felsenstein) model of sequence evolution. Ultrafast bootstrap with 1,000 replicates under iteration of 500 and correlation coefficient of 0.9 are used to infer the ML tree.

Then, based on BEAST 1.10.4, a lognormal distribution with an uncorrelated relaxed clock model was run by using the GTR + F + I site model with four gamma categories, with a random starting tree and a Yule speciation process tree prior (*Suchard et al., 2018*). MCMC was performed with 500 million generations and sampling every 50,000 generations and the effective sample size (ESS) values were confirmed exceeded 200 for all parameters. Then we used the phyutility software to generate an all-compatible consensus tree with the first 25% of the trees as burn-in (*Smith & Dunn, 2008*). Node ages of the consensus phylogeny were estimated using the TreeAnnotator software (*Drummond &*

Rambaut, 2007; *Dexter & Chave, 2016*). A total of 95% highest posterior density intervals (HPD) for each node are shown on the tree. Additionally, the phylogeny was calibrated using four fossils, one fossil from a related clade and by setting the split between *Sloanea* and *Vallea* to 55 ± 2 Mya (*Mayr, 2000*). We used the 40 ± 10 Mya split between *Vallea* and *Aristotelia* as the calibration point (*Heibl & Renner, 2012*). *Elaeocarpus* from the Tasmania in Australia that is about 55 ± 2 Mya old (*Hill, 1984*). The other fossils are leaves of *Rourea* (Connaraceae) from Panama, dated to 59 Mya (*Graham, 1988*). The fossils are used as the ages of the nodes of the tree and applied as calibration points with the normal prior distribution. The tree was viewed and edited by FigTree version 1.4.0 software (http://tree.bio.ed.ac.uk/software/figtree/).

RESULTS

Overall structure

The 11 sequenced Elaeocarpaceae chloroplast genomes showed a quadripartite structure, an LSC region, an SSC region, and a pair of IR regions, with lengths ranging from 157,546 bp (*S. sinensis*) to 159,400 bp (*C. patagua*). The length of the LSC region ranged from 85,874 bp (*E. japonicus*) to 88,413 bp (*S. sinensis*), that of the IR regions ranged from 25,984 bp (*S. sinensis*) to 27,437 bp (*E. japonicus* and *E. japonicus* var. *yunnanensis*), and that of the SSC region ranged from 16,981 bp (*E. japonicus*) to 17,958 bp (*C. Patagua*). The total GC content of the 11 chloroplast genomes from five representative genera was approximately 37%, while the GC contents of the IR, LSC and SSC regions were approximately 43%, 35% and 31%, respectively. In contrast to the chloroplast genome of *A. fruticosa*, which had 133 genes, including eight rRNA genes, 37 tRNA genes, and 88 protein-coding genes, the chloroplast genomes of the other four genera had 132 genes, including eight rRNA genes, 37 tRNA genes, and 87 protein-coding genes. A total of 114 unique genes were detected in the chloroplast genome of *A. fruticosa*, while *C. patagua*, *V. stipularis* and the genus *Sloanea* had 113 unique genes, and genus *Elaeocarpus* had 111 unique genes (Table 3).

Chloroplast genome comparisons

The five genera shared 111 protein-coding genes, but *rpl*32 was only detected in the SSC region of the chloroplast genome of *A. fruticosa* (Fig. 1). The LSC region of the chloroplast genomes lacked *ndh*K in the genus *Elaeocarpus*, *V. stipularis*, and *A. fruticosa* and lacked *inf*A in the genus *Elaeocarpus* and *C. patagua*. *ycf*68 was found in *V. stipularis*, *A. fruticosa* and *C. patagua*. In addition, synteny was detected in the five genera of Elaeocarpaceae (Fig. 2). A significant degree of synteny was found between *V. stipularis* and *A. fruticosa*, *Elaeocarpus* and *Sloanea*. However, the synteny between *C. patagua* and the other four genera was low. Five genera of Elaeocarpaceae were compared, in addition to the species in *Elaeocarpus* and *Sloanea*. Two groups, *E. angustifolius* and *E. hainanensis* as well as *E. japonicus* and *E. japonicus* var. *yunnanensis*, had more blocks of synteny in the genus *Elaeocarpus*. Several blocks of synteny were detected in the four chloroplast genomes of the genus *Sloanea*, suggesting that the four species are similar to each other.

Table 3 Summary of 11 complete chloroplast genomes of Elaeocarpaceae.

	Aristotelia fruticosa	*Crinodendron patagua*	*Vallea stipularis*	*Elaeocarpus angustifolius*	*Elaeocarpus hainanensis*	*Elaeocarpus japonicus*	*Elaeocarpus japonicus* var. *yunnanensis*	*Sloanea sinensis*	*Sloanea cordifolia*	*Sloanea dasycarpa*	*Sloanea longiaculeatae*
Total cpDNA size (bp)	158,085	159,400	158,456	158,315	157,562	157,639	158,124	157,546	158,059	157,966	157,918
Length of the LSC region (bp)	87,427	88,036	87,495	86,465	85,967	85,784	85,928	87,903	88,413	88,297	88,284
Length of the IR regions (bp)	26,477	26,703	26,615	27,038	27,135	27,437	27,437	25,984	25,985	26,011	25,985
Length of the SSC region (bp)	17,704	17,958	17,731	17,774	17,325	16,981	17,322	17,675	17,676	17,647	17,664
Total GC content	37.0%	37.0%	37.0%	36.9%	37.1%	37.1%	37.1%	37.3%	37.2%	37.2%	37.2%
GC content of the IR regions/%	42.5%	42.7%	42.4%	42.3%	42.3%	42.2%	42.2%	42.9%	42.9%	42.9%	42.9%
GC content of the LSC region/%	34.9%	34.7%	34.9%	34.8%	34.9%	35.0%	34.9%	35.1%	35.0%	35.1%	35.0%
GC content of the SSC region/%	30.9%	30.8%	30.9%	31.0%	31.2%	31.3%	31.2%	31.4%	31.3%	31.4%	31.3%
Total number of genes (unique)	133 (114)	132 (113)	132 (113)	132 (111)	132 (111)	132 (111)	132 (111)	132 (113)	132 (113)	132 (113)	132 (113)
Protein-encoding genes	88	87	87	87	87	87	87	87	87	87	87
tRNAs	37	37	37	37	37	37	37	37	37	37	37
rRNAs	8	8	8	8	8	8	8	8	8	8	8

IR expansion and contraction

In the sequenced chloroplast genomes of Elaeocarpaceae, two complete or fragmented copies of *rps*19 and *rpl*2 were located at the boundaries between the LSC region and IRa or IRb region in *V. stipularis*, *A. fruticosa*, *C. patagua* and the genus *Sloanea* (Fig. 3). In contrast, *rps*3, *rpl*22 and *rpl*16 were detected in the neighboring regions of the LSC or IRa or IRb region in the genus *Elaeocarpus*. The distance between the fragment of *ndh*F and the boundary of the SSC and IRb regions in *E. angustifolius* was 370 bp, much greater than that in the chloroplast genomes of other species in the genus *Elaeocarpus*: *E. japonicus* var. *yunnanensis*, *E. japonicus* and *E. angustifolius*. Moreover, the lengths of *ndh*F and *ycf*1 in *E. angustifolius* were shorter than those in the other three species. For the genus *Sloanea*,

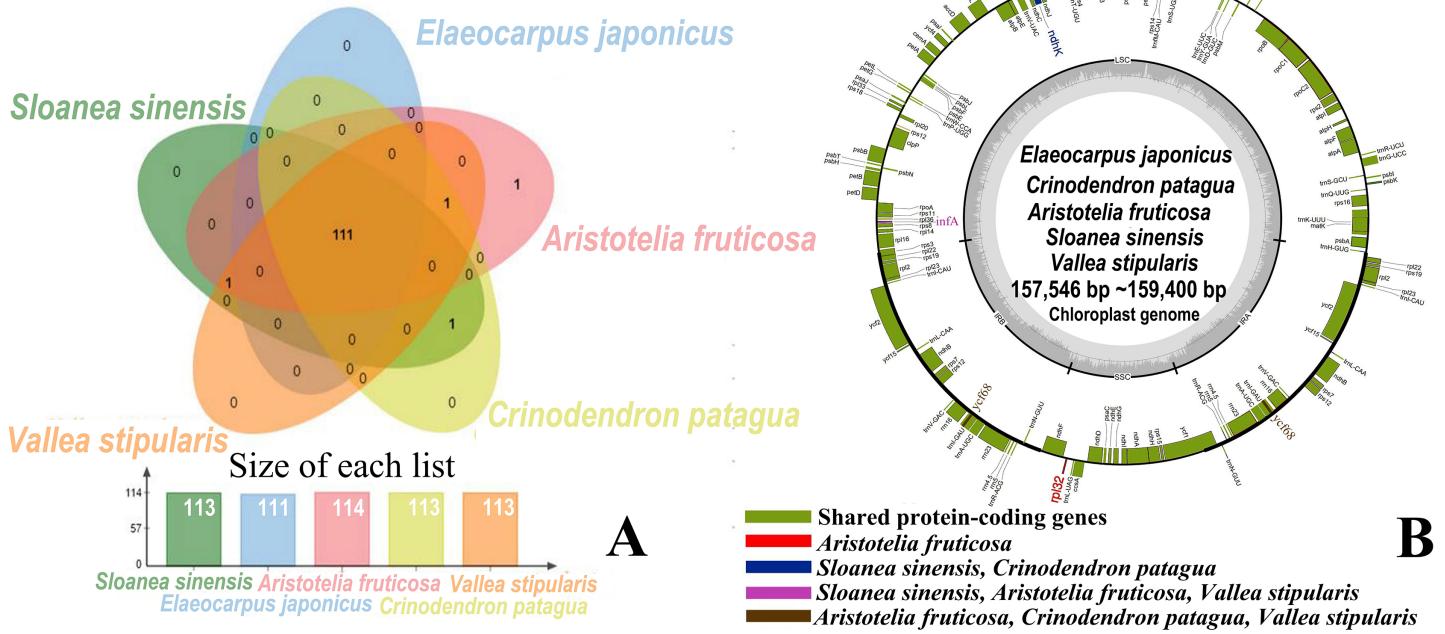



Figure 1 **Shared protein-coding genes in Elaeocarpaceae chloroplast genomes.** (A) Shared protein-coding genes in Elaeocarpaceae chloroplast genomes. The Venn diagram illustrates the number of genes shared between the chloroplast genomes of *Aristotelia fruticosa*, *Crinodendron patagua*, *Vallea stipularis*, *Elaeocarpus japonicus* and *Sloanea sinensis*. (B) Chloroplast genome map of *Aristotelia fruticosa*, *Crinodendron patagua*, *Vallea stipularis*, *Elaeocarpus japonicus* and *Sloanea sinensis*. The green block represents shared protein-coding genes. The red block represents the genes unique to *Aristotelia fruticosa*. The blue block represents the genes unique to *Sloanea sinensis* and *Crinodendron patagua*. The pink block represents the genes unique to *Sloanea sinensis*, *Aristotelia fruticosa* and *Vallea stipularis*. The brown block represents the genes unique to *Aristotelia fruticosa*, *Crinodendron patagua* and *Vallea stipularis*.

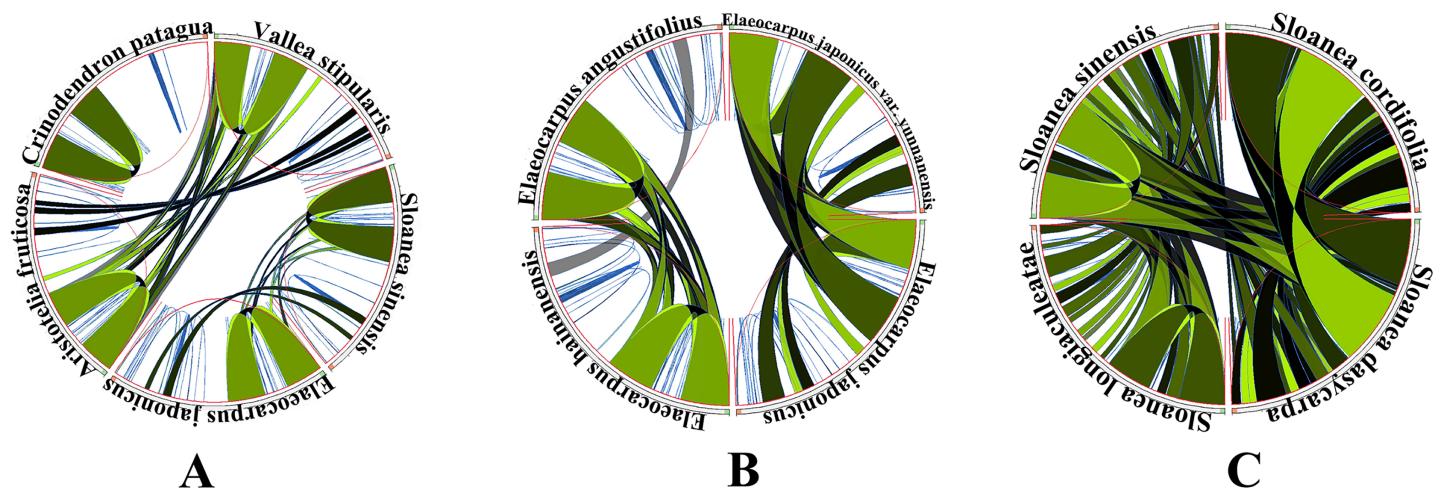
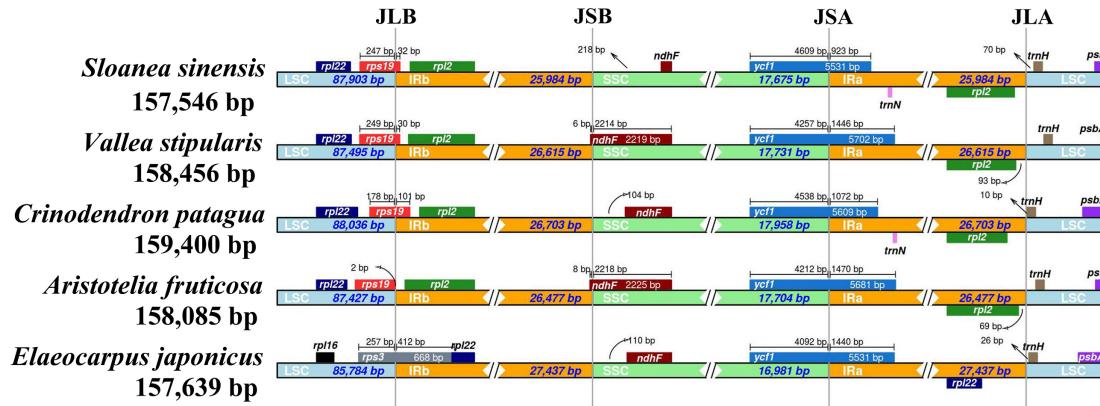
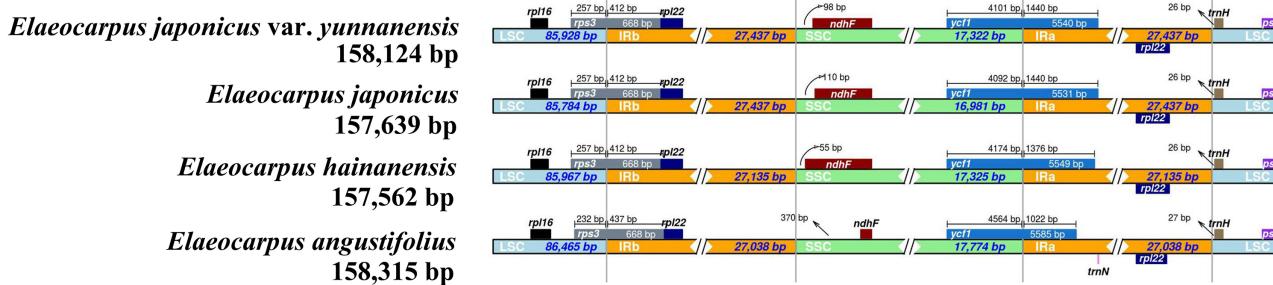



Figure 2 **Synteny detected in Elaeocarpaceae using Circoletto.** (A) Synteny detected between the chloroplast genomes of the Elaeocarpaceae species *Aristotelia fruticosa*, *Crinodendron patagua*, *Vallea stipularis*, *Elaeocarpus japonicus* and *Sloanea sinensis* using Circoletto. (B) Synteny detected between the chloroplast genomes of the Elaeocarpaceae species *Elaeocarpus angustifolius*, *Elaeocarpus japonicus*, *Elaeocarpus japonicus* var. *yunnanensis* and *Elaeocarpus hainanensis* using Circoletto. (C) Synteny detected between the chloroplast genomes of the Elaeocarpaceae species *Sloanea cordifolia, Sloanea dasycarpa, Sloanea longiaculeatae* and *Sloanea sinensis* using Circoletto.

Inverted Repeats

A

Inverted Repeats

B

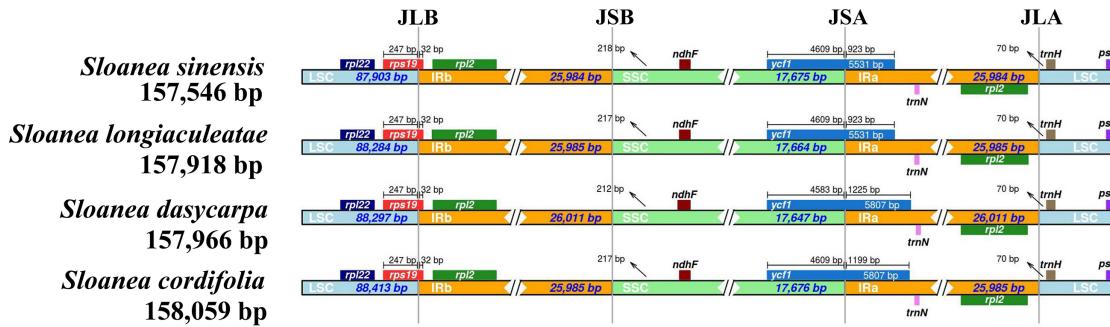
Inverted Repeats

C

Figure 3 Comparisons of IR expansion and contraction in Elaeocarpaceae. (A) The chloroplast genome boundaries of *Aristotelia fruticosa*, *Crinodendron patagua*, *Vallea stipularis*, *Elaeocarpus japonicus* and *Sloanea sinensis* of Elaeocarpaceae. (B) The chloroplast genome boundaries of *Elaeocarpus angustifolius*, *Elaeocarpus japonicus*, *Elaeocarpus japonicus* var. *yunnanensis* and *Elaeocarpus hainanensis*. (C) The chloroplast genome boundaries of *Sloanea cordifolia*, *Sloanea dasycarpa*, *Sloanea longiaculeatae* and *Sloanea sinensis*.

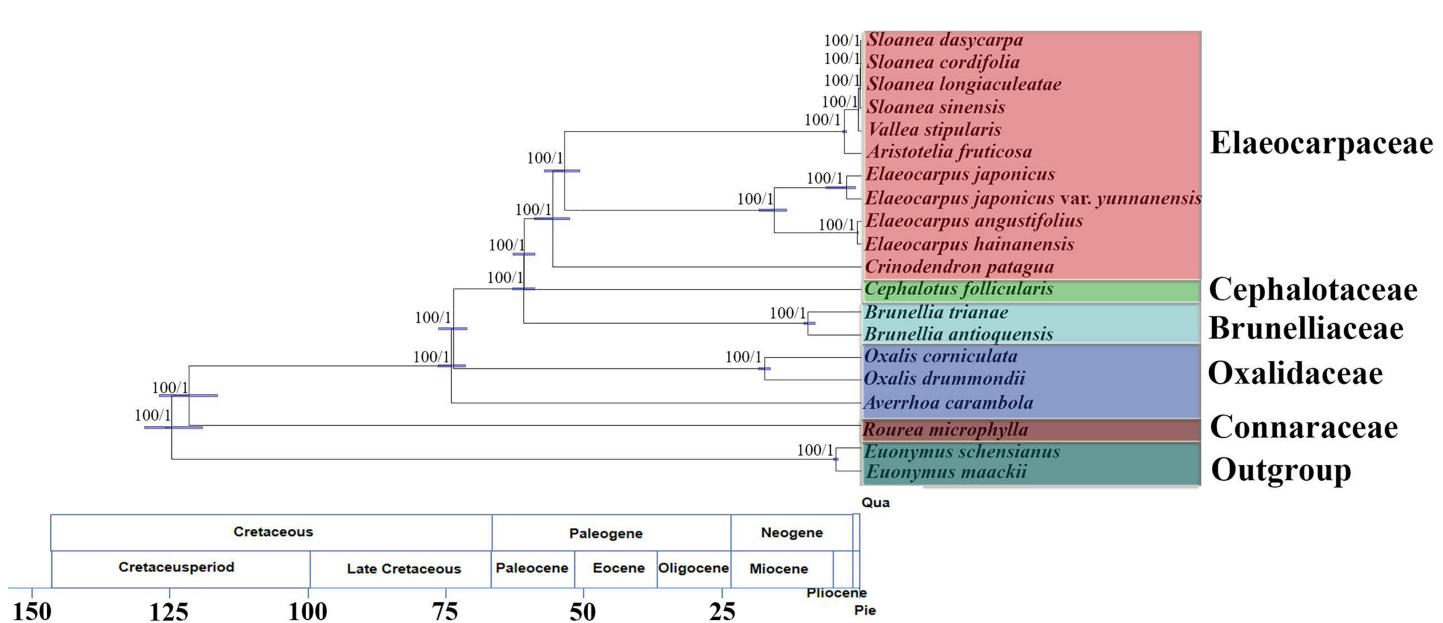



Figure 4 Molecular phylogenomic tree of 20 species of Oxalidales. Molecular phylogenomic tree of 20 species of Oxalidales based on complete chloroplast genome sequences constructed using Bayesian Inference (BI) and Maximum Likelihood (ML). Numbers at each node are bootstrap support values and posterior probability. The blue bars indicate 95% highest posterior density intervals (HPD) of the age estimate. Divergence time of clades and subclades are displayed on the bottom of picture.

the chloroplast genomes of four species, *S. sinensis*, *S. cordifolia*, *S. dasycarpa* and *S. longiaculeatae*, were generally the same in terms of IR expansion and contraction, with the exception that the length of *ycf*1 in *S. dasycarpa* and *S. longiaculeatae* was greater than that in *S. sinensis* and *S. cordifolia*.

Phylogenomics and molecular clock dating analysis

The matrix of complete chloroplast genomes was used to reconstruct a phylogenomic tree of Oxalidales (Fig. 4). The phylogenetic tree constructed by the ML analysis is consistent with the phylogenetic tree based on BI method. The phylogenies show high robustness with the highest posterior probability and bootstrap value for all clades. The molecular tree showed that *R. microphylla* representing Connaraceae started to diversify at 119 Mya (95% highest posterior probability density intervals (HPD):114–125 Mya). *Averrhoa carambola*, *Oxalis drummondii* and *Ox. corniculata* representing Oxalidaceae diverged from basal Oxalidales (ca. 122 Mya) at 73 Mya (HPD:70–75 Mya). Cephalotaceae (ca. 60 Mya; HPD:58–62 Mya) has the closest genetic relationship with Elaeocarpaceae (ca. 60 Mya; HPD:58–62 Mya) with that of Oxalidaceae (ca. 119 Mya; HPD:114–125 Mya). Brunelliaceae has a similar differentiation time about 60 Mya (HPD:58–62 Mya) from Oxalidaceae.

The molecular tree also showed the sister relationships of 11 chloroplast genomes from five representative genera of Elaeocarpaceae were highly supported. Clade I, containing the genus *Elaeocarpus*, was 100% supported and was dated to ca. 53 Mya (HPD:50–56 Mya), and the crown node age of clade II (*C. patagua*) was dated to ca. 55 Mya (HPD:52–58 Mya). Diversification of clade III, containing the *Sloanea* alliance (*V. stipularis*, *A. fruticosa*

and the genus *Sloanea*), was dated to 53 Mya (HPD:50–56 Mya). Further differentiation of *V. stipularis* and *A. fruticosa* took place within the last 3 Mya (HPD:2–3 Mya). In addition, the genus *Sloanea* started to diversify during the late Miocene (ca. 0.4 Mya; HPD:0.3–0.4 Mya).

DISCUSSION

Complete chloroplast structure of Elaeocarpaceae

This study included 11 complete chloroplast genomes for Elaeocarpaceae plants. All these complete chloroplast genomes had a total GC content of 37%, consistent with the low GC content in the chloroplast genomes of other angiosperms. The higher the content of GC is, the higher the density of DNA and the more conserved the chloroplast genome is *Do, Kim & Kim (2013)*. Therefore, variation might occur in the SSC region rather than the IR regions. Comparisons of the 11 plastomes showed the loss of *inf*A in *C. patagua* and the genus *Elaeocarpus*, and similar losses or pseudogenization was reported in the 309 complete chloroplast genomes of 24 species of angiosperms (*Millen et al., 2001*). *ndh* genes are frequently pseudogenized or lost in plant groups with a degree of heterotrophy due to evolutionary adaptation to excessive water in the environment, as observed in *Aneura*, *Cuscuta*, *Epifagus*, *Hydnora*, and nonphotosynthetic orchid species and some autotrophic gymnosperms and ferns (*DePamphilis & Palmer, 1990*; *McNeal et al., 2007*; *Wickett et al., 2008*; *Wicke et al., 2011*; *Kim et al., 2015*; *Naumann et al., 2016*). The *rpl*32 gene was detected in *A. fruticosa* but not in the other four genera (*V. stipularis, C. patagua*, the genus *Elaeocarpus* and the genus *Sloanea*), which is similar to previously published research about the losses of two genes, *inf*A and *rpl*32, in *Thalictrum coreanum* (*Park & Jansen, 2015*). In summary, the five genera may have experienced different niche expansions.

Previous studies have shown that IR boundary regions with large expansions and contractions may be related to double-strand breakage and repair, while small expansions and contractions may be related to gene conversions, which is a common phenomenon in the evolution of the chloroplast genomes (*Kim & Lee, 2004*; *Khakhlova & Bock, 2006*; *Hansen et al., 2007*; *Wang et al., 2008*; *Ma et al., 2013*; *Liang, Wen & Gao, 2018*). We found large IR expansions in the five genera. The genus *Elaeocarpus* is different from the other four genera at the IR/SC boundary, which may reflect that the genus *Sloanea* has an older origin and experienced a different evolution event. In addition, *rps*19 was located across the LSC/IRB regions in four genera, while the boundary of the LSC and IRb regions in the genus *Elaeocarpus* included *rps*3. Research shows that the locations of *rps*19 and *rps*3 differ between the chloroplasts of monocotyledons and dicotyledons. In some dicotyledons, *rps*19 only partially exists in the IR region, while the *rps*3 gene is only found in *Paris* and Melanthiaceae (*Lin et al., 2012*; *Sarah, Kim & Kim, 2013*). Compared with the other four genera, the genus *Sloanea* experienced different complex evolutionary events.

Homologous fragments have been found *via* collinearity analysis in various plants, including Capparaceae (*Alzahrani et al., 2021*), Ranunculaceae (*Park & Park, 2021*), and *Passiflora* (*Cauz-Santos et al., 2020*). The length of homologous fragments is related to the time of divergence between species. The shorter the time of species differentiation is, the more homologous fragments there are (*Cheng et al., 2013*). According to the similarity of the 11 chloroplast genomes of Elaeocarpaceae, we detected several blocks of synteny

between *V. stipularis* and *A. fruticosa*, the genus *Elaeocarpus* and the genus *Sloanea*, meaning that the times of divergence between the genus *Sloanea* and the genus *Elaeocarpus*, *V. stipularis* and *A. fruticosa* was similar. Interestingly, there were no blocks of synteny in *C. patagua* with the other four genera, meaning that the evolution of *C. patagua* was different from the rest of the genera. In the genus *Elaeocarpus* and genus *Sloanea*, it is worth noting that the divergence time of *E. japonicus* was similar to that of *E. japonicus* var. *yunnanensis* and that of *E. angustifolius* was similar to that of *E. hainanensis*. In addition, the times of divergence among species in the genus *Sloanea* were similar.

Phylogenomic relationships and historical biogeography in Oxalidales

Based on the 20 species of six families with available complete chloroplast genomes, a phylogenomic tree of Oxalidales was reconstructed, consistent with the recent phylogeny (*Byng et al., 2016*; *Baker et al., 2021*; *Li et al., 2021*). Elaeocarpacea was clarified as sister to Cephalotaceae and Brunelliaceae and Connaraceae. In addition, Oxalidaceae is far from Elaeocarpaceae as previously recognized, which was recognized by *Heibl & Renner (2012)*.

Pillon et al. (2021) phylogeny of Oxalidales based on plastid genomics has been used as data for event-based biogeographic analysis of the world. In that study the likely ancestral area for the Oxalidales is Australia/New Guinea + New Caledonia in Cretaceous, which was consistent with our results. Indeed, the greatest number of extant species and genera are found in Oceania, and particularly in eastern Australia, New Guinea, and New Caledonia (*Kershaw, 1976*; *Kershaw, Bretherton & van der Kaars, 2007*; *Sniderman, 2011*).

The age of the Connaraceae clade with *R. microphylla* was much older than previously estimated (74 Mya; *Heibl & Renner, 2012*). The recent discovery of *Connarus*-like wood from the Paleocene of India, outside the modern range of the family, suggests a possible origin in India during the Cretaceous, when India was an island continent, and subsequently spread throughout the Old World tropics as India docked with Asia (*Baas et al., 2017*).

The differentiation time of Oxalidaceae is consistent with that of Heibl and Renner, which is about 68 Mya. Geographical distribution patterns suggest the origin of the family in the southern hemisphere, prior to the separation of South America and Africa (*Raven & Axelrod, 1974*).

The split from Cephalotaceae and Brunelliaceae was estimated at 60 Mya, more recent than Heibl and Renner's research (78 Mya). According to recent research, *Brunellia* is exclusively American, with only six of the 61 known species occurring north of Panama. Cephalotaceae grows only in the extreme SouthWest of Australia (*Matthews & Endress, 2002*). The presence of Brunelliaceae and Cephalotaceae may indicate that the genera may have been represented north of Panama before the closing of the central American land bridge (*Coode, 2004*).

As for the differentiation time of Elaeocarpaceae, it has long been postulated that the family Elaeocarpaceae originated in the southern hemisphere, of which only *Elaeocarpus* and *Sloanea* reach the northern hemisphere (*Raven & Axelrod, 1974*). The crown age of Elaeocarpaceae estimated in this study based on molecular data was younger than the age

previously estimated at 79.62–85.2 Mya (*Magallon et al., 2015*; *Phoon, 2015*), 64–66 Mya (*Wikström, Savolainen & Chase, 2001*), 67 Mya (*Heibl & Renner, 2012*) and 100 Mya (*Crayn, Rossetto & Maynard, 2006*), older than 38 Mya (*Harris & Davies, 2016*). These differences may be due to the choice of DNA markers and the accuracy of the fossil calibrations of molecular evolutionary rates. The earliest divergence within the Elaeocarpaceae appears to have occurred in the late Cretaceous based on our data, which is broadly coincident with the time when the western (Africa and South America) and eastern (Australia, Antarctica, Madagascar and India) parts of Gondwana were separating (*Ali & Aitchison, 2008*).

Phylogenomic relationships and historical biogeography in Elaeocarpaceae

Within Elaeocarpaceae, the 11 taxa were separated into the following groups in our study: the *Sloanea* alliance (*V. stipularis*, *A. fruticosa* and *Sloanea*), *Elaeocarpus* alliance and *C. patagua* alliance. The phylogenomic placements are consistent with those in Phoon's research (*Phoon, 2015*). One major challenge in previous studies of the phylogenetic relationships between and within Elaeocarpaceae was the focus on DNA markers (*trn*L-*trn*F region and *trn*V-*ndh*C region) rather than complete chloroplast genomes (*Maynard, 2004*; *Baba, 2013*; *Phoon, 2015*). Furthermore, the DNA markers exhibited low sequence variability, leading to insufficiently resolved phylogenies within *Elaeocarpus*, and there no phylogenetic tree was constructed for *Sloanea*. Our phylogenomic analysis strongly confirmed the preliminary results of previous studies but with higher robustness, and the results of the analysis improved the posterior probabilities of all clades (*Maynard, 2004*; *Baba, 2013*; *Phoon, 2015*).

Compared with the differentiation of *V. stipularis* and *A. fruticosa* in Phoon's study, the age of the split between *V. stipularis* and *A. fruticosa* was much younger than previously estimated at 37 Mya (*Phoon, 2015*). The results of the present study agreed with Coode's phylogenetic reconstruction in which *Vallea* and *A.* were sisters and Phoon's finding that the ancestors may have dispersed between western and eastern Gondwana. The minimum estimates of divergence times between *V. stipularis* and *A. fruticosa* because the divergence of the South American and New Zealand lineages at 24–27 and 3 Mya respectively, postdates the isolation of their respective landmasses (*McLoughlin, 2001*).

Crinodendron was resolved in this study as an independent branch. The split from Elaeocarpaceae was estimated at 55 Mya, more recent than Phoon's estimate (59 Mya). The divergence of *Crinodendron* is estimated to have occurred during the Paleo-Eocene, but the origin of the genus is almost certainly older given the position of *Dubouzetia brasiliense* (from a dwarf cloud forest near the Atlantic coast of Brazil) as sister to the rest of the genus, based on morphological data (*Coode, 2004*).

Elaeocarpus represents a widespread lineage in Elaeocarpaceae that diverged 53 Mya, which was older than Phoon's estimate (40 Mya). Divergence time analysis suggests that *Elaeocarpus* split in the Eocene and migrated out of Australia to the surrounding regions mostly in the Oligocene and the Miocene, although the taxon sampling without species of

Southeast Asian in this clade has led to these dates being doubted (*Crayn, Rossetto & Maynard, 2006*; *Heibl & Renner, 2012*; *Phoon, 2015*).

Divergence time analysis suggests that *Sloanea* diverged from its sister species *V. stipularis* and *A. fruticosa* at 0.4 Mya, more recent than Phoon's estimate (29 Mya). The reason may be that *Sloanea* in China belongs to a branch of the *Slonaea* genus, and the differentiation time is late.

Overall, the divergence times of all genera in Elaeocarpaceae inferred using the complete chloroplast genomes were more accurate than those inferred using DNA markers (*trn*L-*trn*F region and *trn*V-*ndh*C region).

CONCLUSIONS

In the present research, the chloroplast genomes of nine species were assembled using Illumina high-throughput sequencing data. The genomic structures of the 11 samples were compared and analyzed. On this basis, we concluded that the chloroplast genome structure and gene size in Elaeocarpaceae showed some difference. In addition, we determined the relationships of Elaeocarpaceae within Oxalidales and inferred divergence dates and the biogeographic history of the major clades within Oxalidales and Elaeocarpaceae. These results are consistent with previous studies reporting relationship in Elaeocarpaceae and provide new insight into the evolution of the Elaeocarpaceae.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr. Hai-Zhong Yu and Prof. Ren-Lin Liu from Gannan Normal University.

Funding

This work was supported by the National Natural Science Foundation of China (Grant numbers 31110103911, J1310002), and the Science and Technology Project in Jiangxi Province Department of Education (GJJ170839). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 31110103911, J1310002.
Science and Technology Project in Jiangxi Province Department of Education: GJJ170839.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yihui Wang performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Yifei Xie conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Jiayi Jin performed the experiments, prepared figures and/or tables, and approved the final draft.
- Jinyue Li performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xiangdong Qiu performed the experiments, prepared figures and/or tables, and approved the final draft.
- Yang Tong performed the experiments, prepared figures and/or tables, and approved the final draft.
- Zhongyang Li conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Zhixiang Zhang conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Wenling Lai analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The data are available at GenBank: MT683335, MW004670, MT982369, MW602804, MW242791, MW242788, MW242787, MW242790, MT982370, MW242790, MT982368.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.15322#supplemental-information.

REFERENCES

Ali JR, Aitchison JC. 2008. Gondwana to Asia: plate tectonics, paleogeography and the biological connectivity of the Indian subcontinent from the Middle Jurassic through end Eocene (166–35 Ma). *Earth-Science Reviews* **88(3–4)**:145–166 DOI 10.1016/j.earscirev.2008.01.007.

Ali A, Jaakko H, Peter P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34(17)**:3030–3031 DOI 10.1093/bioinformatics/bty220.

Alzahrani DA, Enas JA, Samaila SY, Abba A. 2021. Comparative analysis of chloroplast genomes of four medicinal capparaceae species: genome structures, phylogenetic relationships and adaptive evolution. *Plants* **10(6)**:1229 DOI 10.3390/plants10061229.

Anthony JM, Allcock RJ, Krauss SL. 2016. Isolation and characterization of 13 microsatellites for the rare endemic shrub *Tetratheca erubescens* (Elaeocarpaceae). *Applications in Plant Sciences* **4(2)**:1500102–1500105 DOI 10.3732/apps.1500102.

Baas P, Manchester S, Wheeler EA, Srivastava R. 2017. Fossil wood with dimorphic fibers from the Deccan Intertrappean Beds of India-the oldest fossil Connaraceae? *IAWA Journal* **38(1)**:124–133 DOI 10.1163/22941932-20170162.

Baba Y. 2013. Evolution, systematics, and taxonomy of *Elaeocarpus* (Elaeocarpaceae) in Australasia. PhD thesis, James Cook University, Townsville, Australia.

Baker WJ, Bailey P, Barber V, Barker A, Bellot S, Bishop D, Botigué LR, Brewer G, Carruthers T, Clarkson JJ, Cook J, Cowan RS, Dodsworth S, Epitawalage N, Françoso E, Gallego B, Johnson MG, Kim JT, Leempoel K, Maurin O, McGinnie C, Pokorny L, Roy S, Stone M, Toledo E, Wickett NJ, Zuntini AR, Eiserhardt WL, Kersey PJ, Leitch IJ, Forest FA. 2021. Comprehensive phylogenomic platform for exploring the angiosperm tree of life. *Systematic Biology* **71(2)**:301–319 DOI 10.1093/sysbio/syab035.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19(5)**:455–477 DOI 10.1089/cmb.2012.0021.

Barrett CF, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, Santos C. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* **31(12)**:3095–3112 DOI 10.1093/molbev/msu252.

Bock R. 2007. Plastid biotechnology: prospects for herbicide and insect resistance, metabolic engineering, and molecular farming. *Current Opinion in Biotechnology* **18(2)**:100–106 DOI 10.1016/j.copbio.2006.12.001.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30(15)**:2114–21120 DOI 10.1093/bioinformatics/btu170.

Braukmann TWA, Kuzmina M, Stefanovic S. 2009. Loss of all plastid *ndh* genes in Gnetales and conifers: extent and evolutionary significance for the seed plant phylogeny. *Current Genetics* **55(3)**:323–337 DOI 10.1007/s00294-009-0249-7.

Byng JW, Chase MW, Maarten JM, Christenhusz MJM, Fay MF, Judd WS, Soltis DE, Mabberley DJA, Sennikov N, Soltis PS, Stevens PF, Barbara B, Brockington S, Chautems A, Clark J, Conran J, Haston E, Möller M, Moore M, Olmstead R, Perret M, Skog L, Smith J, Tank D, Vorontsova M, Weber A. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181(1)**:1–20 DOI 10.1111/boj.12385.

Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK. 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. *BMC Evolutionary Biology* **6(1)**:77–97 DOI 10.1186/1471-2148-6-77.

Cauz-Santos LA, Da-Costa ZP, Callot C, Cauet S, Zucchi MI, Bergès H, Berg CVD, Lucia M, Vieira C. 2020. A repertory of rearrangements and the loss of an inverted repeat region in Passiflora chloroplast genomes. *Genome Biology and Evolution* **12(10)**:1841–1857 DOI 10.1093/gbe/evaa155.

Cheng F, Mandakova T, Wu J, Xie Q, Lysak MA, Wang X. 2013. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. *The Plant Cell Online* **25(5)**:1541–1554 DOI 10.1105/tpc.113.110486.

Christenhusz M, Byng JW. 2016. The number of known plant species in the world and its annual increase. *Phytotaxa* **261(3)**:201–217 DOI 10.11646/phytotaxa.261.3.1.

Coode M. 2004. Elaeocarpaceae. In: Kubitzki K, ed. *Flowering Plants, Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales.* Vol. 6. Berlin, Heidelberg: Springer-Verlag, 135–143.

Crayn DM, Rossetto M, Maynard DJ. 2006. Molecular phylogeny and dating reveal an Oligo-Miocene radiation of dry-adapted shrubs (Tremandraceae) from rainforest tree progenitors (Elaeocarpaceae) in Australia. *American Journal of Botany* **93(9)**:1328–1342 DOI 10.3732/ajb.93.9.1328.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics, and parallel computing. *Nature Methods* **9(8)**:772 DOI 10.1038/nmeth.2109.

Darzentas N. 2010. Circoletto: visualizing sequence similarity with Circos. *Bioinformatics* **26(20)**:2620–2622 DOI 10.1093/bioinformatics/btq484.

DePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **348(6299)**:337–339 DOI 10.1038/348337a0.

Dexter K, Chave J. 2016. Evolutionary patterns of range size, abundance and species richness in Amazonian angiosperm trees. *PeerJ* **4**:e2402–e2416 DOI 10.7717/peerj.2402.

Do HDK, Kim JS, Kim JH. 2013. Comparative genomics of four Liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). *Gene* **530(2)**:229–235 DOI 10.1016/j.gene.2013.07.100.

Doyle JJ, Doyle JL, Palmer JD. 1995. Multiple independent losses of 2 genes and one intron from legume chloroplast genomes. *Systematic Botany* **20(3)**:272–294 DOI 10.2307/2419496.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **8(7)**:214–222 DOI 10.1186/1471-2148-7-214.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Research* **32(Web Server issue)**:W273–W280 DOI 10.1093/nar/gkh458.

Gao L, Su YJ, Wang T. 2010. Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *Journal of Systematics and Evolution* **48(2)**:77–93 DOI 10.1111/j.1759-6831.2010.00071.x.

Goremykin V, Hirsch-Ernst K, Wölfl S, Hellwig FH. 2003. The chloroplast genome of the basal angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. *Plant Systematics and Evolution* **242(1)**:119–135 DOI 10.1007/s00606-003-0056-4.

Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH. 2005. Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. *Molecular Biology and Evolution* **22(9)**:1813–1822 DOI 10.1093/molbev/msi173.

Graham A. 1988. Studies in neotropical paleobotany. V. The lower miocene communities of panama-the culebra formation. *Annals of the Missouri Botanical Garden* **75(4)**:1440–1466 DOI 10.2307/2399295.

Guindon S, Gascuel O, Rannala B. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52(5)**:696–704 DOI 10.1080/10635150390235520.

Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK. 2010. Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *Journal of Molecular Evolution* **70(2)**:149–166 DOI 10.1007/s00239-009-9317-3.

Hansen DR, Dastidar SG, Cai Z, Penaflor C, Kuehl JV, Boore JL, Jansen RK. 2007. Phylogenetic and evolutionary implications of complete chloroplast genome sequences of four early-diverging angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae), *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). *Molecular Phylogenetics and Evolution* **45(2)**:547–563 DOI 10.1016/j.ympev.2007.06.004.

Harris LW, Davies TJ. 2016. A complete fossil-calibrated phylogeny of seed plant families as a tool for comparative analyses: testing the 'time for speciation' hypothesis. *PLOS ONE* **11(10)**:e0162907–e0162923 DOI 10.1371/journal.pone.0162907.

Heibl C, Renner SS. 2012. Distribution models and a date phylogeny for chilean *Oxalis* species reveal occupation of new habits by different lineages, not rapid adaptive radiation. *Systematic Biology* **61(5)**:823–834 DOI 10.1093/sysbio/sys034.

Hill RS. 1984. Tertiary Nothofagus macrofossils from Cethana, Tasmania. *Alcheringa: An Australasian Journal of Palaeontology* **8(1)**:81–86 DOI 10.1080/03115518408619610.

Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21(1)**:241–272 DOI 10.1186/s13059-020-02154-5.

Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KD, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD, Olmstead RG, Renzaglia KS, Wolf PG. 2010. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evolutionary Biology* **10(1)**:321–337 DOI 10.1186/1471-2148-10-321.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30(4)**:772–780 DOI 10.1093/molbev/mst010.

Kershaw AP. 1976. A late Pleistocene and Holocene pollen diagram from Lynch's Crater, north-eastern Queensland, Australia. *New Phytologist* **77(2)**:469–498 DOI 10.1111/j.14698137.1976.tb01534.x.

Kershaw AP, Bretherton SCL, van der Kaars WA. 2007. A complete pollen record of the last 230 ka from Lynch's Crater, north-eastern Australia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **251(1)**:23–45 DOI 10.1016/j.palaeo.2007.02.015.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal* **46(1)**:85–94 DOI 10.1111/j.1365-313X.2006.02673.x.

Kim HT, Kim JS, Moore MJ, Neubig KM, Williams NH, Whitten WM, Kim JH. 2015. Seven new complete plastome sequences reveal rampant independent loss of the *ndh* gene family across orchids and associated instability of the inverted repeat/small single-copy region boundaries. *PLOS ONE* **10(11)**:e0142215–e0142233 DOI 10.1371/journal.pone.0142215.

Kim KJ, Lee HL. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* **11(4)**:247–261 DOI 10.1093/dnares/11.4.247.

Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, Jin JJ, Zeng CX, Wang H, Yu WB, Zhang R, van der Bank M, Olmstead RG, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Yi TS, Li DZ. 2021. Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biology* **19(1)**:232–245 DOI 10.1186/s12915-021-01166-2.

Liang FP, Wen XN, Gao HY. 2018. Analysis of chloroplast genomes features of Asteraceae species. *Genomics and Applied Biology* **37(12)**:5437–5447 DOI 10.1007/s11105-015-0853-2.

Lin CP, Wu CS, Huang YY, Chaw SM. 2012. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biology and Evolution* **4(3)**:374–381 DOI 10.1093/gbe/evs102.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. Organellar genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* **41(Web Server issue))**:W575–W581 DOI 10.1093/nar/gkt289.

Ma J, Yang B, Zhu W, Sun L, Tian J, Wang X. 2013. The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* **528(2)**:120–131 DOI 10.1016/j.gene.2013.07.037.

Magallon S, Gómez AS, Sánchez R, Luna L, Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207(2)**:437–453 DOI 10.1111/nph.13264.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17(1)**:10–12 DOI 10.14806/ej.17.1.200.

Matthews ML, Endress PK. 2002. Comparative structure and systematics in Oxalidales (Oxalidaceae, Connaraceae, Brunelliaceae, Cephalotaceae, Cunoniaceae, Elaeocarpaceae, Tremandraceae). *Botanical Journal of the Linnean Society* **140(4)**:321–381 DOI 10.1046/j.1095-8339.2002.00105.x.

Maynard DJ. 2004. A molecular phylogeny for Australian *Elaeocarpus* (Elaeocarpaceae) and the affinities of a putative new taxon. BSc Thesis, University of New South Wales, Sydney.

Mayr G. 2000. A new mousebird (Coliiformes: Coliidae) from the Oligocene of Germany. *Journal of Ornithology* **141(1)**:85–92 DOI 10.1007/BF01651775.

McCoy SR, Kuehl JV, Boore JL, Raubeson LA. 2008. The complete plastid genome sequence of *Welwitschia mirabilis*: an unusually compact plastome with accelerated divergence rates. *BMC Ecology and Evolution* **8(1)**:130–146 DOI 10.1186/1471-2148-8-130.

McLoughlin S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* **49(3)**:271–300 DOI 10.1071/BT00023.

McNeal JR, Kuehl JV, Boore JL, DePamphilis CW. 2007. Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology* **7(1)**:57–79 DOI 10.1186/1471-2229-7-57.

McPherson H, Porter C, Rymer PD, Crayn DM, Rossetto M. 2008. Isolation and characterization of polymorphic microsatellite loci from *Tetratheca ericifolia* (Elaeocarpaceae). *Molecular Ecology Resources* **8(4)**:867–869 DOI 10.1111/j.1755-0998.2008.02093.x.

Millen RS, Olmstead RG, Adams KL, Palmer JD, Nga T, Heggie LL, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, Calie PJ, Jermiin LS, Wolfe KH. 2001. Many parallel losses of *inf*A from chloroplast angiosperm evolution with multiple independent transfers to the nucleus. *The Plant Cell* **13(3)**:645–658 DOI 10.1105/tpc.13.3.645.

Naumann J, Der JP, Wafula EK, Jones SS, Wagner ST, Honaas LA, Ralph PE, Bolin JF, Maass E, Neinhuis C, Wanke S, DePamphilis CW. 2016. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biology and Evolution* **8(2)**:345–363 DOI 10.1093/gbe/evv256.

Park S, Jansen RK. 2015. Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl*32 gene to the nucleus in the ancestor of the subfamily Thalictroideae. *BMC Plant Biology* **15(1)**:40–52 DOI 10.1186/s12870-015-0432-6.

Park KT, Park SJ. 2021. Phylogenomic analyses of hepatica species and comparative analyses within *Tribe Anemoneae* (Ranunculaceae). *Frontiers in Plant Science* **12**:728–739 DOI 10.3389/fpls.2021.638580.

Philippe B, Jérôme M, Frédéric E, Djemiel C, Klopp C. 2014. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15(1)**:293–300 DOI 10.1186/1471-2105-15-293.

Phoon SN. 2015. Systematics and biogeography of *Elaeocarpus* (Elaeocarpaceae). PhD thesis, James Cook University, Townsville, Australia.

Pillon Y, Hopkins HCF, Maurin O, Epitawalage N, Bradford J, Rogers ZS, Baker WJ, Forest F. 2021. Phylogenomics and biogeography of Cunoniaceae (Oxalidales) with complete generic sampling and taxonomic realignments. *American Journal of Botany* **108(7)**:1181–1200 DOI 10.1002/ajb2.1688.

Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* **15**:50 DOI 10.1186/s13007-019-0435-7.

Raven P, Axelrod DI. 1974. Angiosperm biogeography and past continental movements. *Annals of the Missouri Botanical Garden* **61(3)**:539–673 DOI 10.2307/2395021.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19(12)**:1572–1574 DOI 10.1093/bioinformatics/btg180.

Roper JM, Hansen SK, Wolf PG, Karol KG. 2007. The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). *American Fern Journal* **97(2)**:95–106 DOI 10.1640/0002-8444(2007)97[95:TCPGSO]2.0.CO;2.

Sarah SS, Kim JH, Kim JS. 2013. Complete chloroplast genome of *Chionographis japonica* (Willd.) Maxim. (Melanthiaceae): comparative genomics and evaluation of universal primers for Liliales. *Plant Molecular Biology Reporter* **31(6)**:1407–1421 DOI 10.1007/s11105-013-0616-x.

Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank M, Cameron KM, Johnson SA, Lledó MD, Pintaud JC, Powell M, Sheanhan MC, Soltis DE, Soltis PS, Weston P, Whitten WM, Wurdack KJ, Chase MW. 2000. Phylogeny of the eudicots: a nearly complete familial analysis based on *rbc*L gene sequences. *Kew Bulletin* **55(2)**:257–309 DOI 10.2307/4115644.

Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* **24(5)**:715–716 DOI 10.1093/bioinformatics/btm619.

Sniderman JMK. 2011. Early Pleistocene vegetation change in upland south-eastern Australia. *Journal of Biogeography* **38(8)**:1456–1470 DOI 10.1111/j.1365-2699.2011.02518.x.

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF, Axtell M, Swensen SM, Prince LM, Kress WJ, Nixon KC, Farris JS. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. *Botanical Journal of the Linnean Society* **133(4)**:381–461 DOI 10.1111/j.1095-8339.2000.tb01588.x.

Song Y, Yu WB, Tan Y, Liu B, Yao X, Jin J, Padmanaba M, Yang JB, Corlett RT. 2017. Evolutionary comparisons of the chloroplast genome in lauraceae and insights into loss events in the magnoliids. *Genome Biology and Evolution* **9(9)**:2354–2364 DOI 10.1093/gbe/evx180.

Stamatakis A, Hoover P, Rougemont JA. 2008. Rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57(5)**:758–771 DOI 10.1080/10635150802429642.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* **4(1)**:vey016–vey021 DOI 10.1093/ve/vey016.

Sugai K, Setsuko S, Nagamitsu T, Murakami N, Kato H, Yoshimaru H. 2013. Genetic differentiation in *Elaeocarpus photiniifolia* (Elaeocarpaceae) associated with geographic distribution and habitat variation in the Bonin (Ogasawara) Islands. *Journal of Plant Research* **126(6)**:763–764 DOI 10.1007/s10265-013-0571-5.

Sugai K, Setsuko S, Uchiyama K, Murakami N, Kato H, Yoshimaru H. 2012. Development of EST-SSR markers for *Elaeocarpus photiniifolia* (Elaeocarpaceae), an endemic taxon of the Bonin Islands. *American Journal of Botany* **99(2)**:e84–e87 DOI 10.3732/ajb.1100391.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56(4)**:564–577 DOI 10.1080/10635150701472164.

Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proceedings of the National Academy of Sciences of the United States of America* **91(21)**:9794–9798 DOI 10.1073/pnas.91.21.9794.

Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology* **8(1)**:36–49 DOI 10.1186/1471-2148-8-36.

Wang W, Messing J. 2011. High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLOS ONE* **6(9)**:e24670–e24679 DOI 10.1371/journal.pone.0024670.

Wang YH, Zhang ZX, Xie YF. 2021. The complete chloroplast genome of *Elaeocarpus japonicus* Sieb. et Zucc. (Elaeocarpaceae). *Mitochondrial DNA Part B-Resources* **6(2)**:557–559 DOI 10.1080/23802359.2021.1872454.

Weng YH, Ye DQ, You YF, Chen YT, Fan FJ, Shi JS, Chen JH. 2021. The complete chloroplast genome sequence of *Sloanea sinensis*. *Mitochondrial DNA Part B-Resources* **6(2)**:555–556 DOI 10.1080/23802359.2021.1872453.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31(20)**:3350–3352 DOI 10.1093/bioinformatics/btv383.

Wicke S, Müller KF, DePamphilis CW, Quandt D. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences of the United States of America* **113(32)**:9045–9050 DOI 10.1073/pnas.1607576113.

Wicke S, Schneeweiss GM, DePamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76(3–5)**:273–297 DOI 10.1007/s11103-011-9762-4.

Wickett NJ, Zhang Y, Hansen SK, Roper JM, Kuehl JV, Plock SA, Wolf PG, DePamphilis CW, Boore JL, Goffinet B. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Molecular Biology and Evolution* **25(2)**:393–401 DOI 10.1093/molbev/msm267.

Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society B: Biological Sciences* **268(1482)**:2211–2220 DOI 10.1098/rspb.2001.1782.

Wu ML. 2016. Phylogeny and chloroplast genome evolution of Bambusoideae. PhD thesis, Fujian Agriculture and Forestry University, Fujian, China.

Wu CS, Lai YT, Lin CP, Wang WN, Chaw SM. 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Molecular Phylogenetics and Evolution* **52(1)**:115–124 DOI 10.1016/j.ympev.2008.12.026.