Comparison of the Chloroplast Genomes and Phylogenomic Analysis of Elaeocarpaceae

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

[1]School of Life Sciences, Gannan Normal University, Ganzhou, P.R. China

[2]Key Laboratory of Nanling Plant Resources Conservation and Utilization, Ganzhou, P.R. China

[3]Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Science, Yunnan, P.R. China

[4]School of Ecology and Nature conservation, Beijing Forestry University, Beijing, P.R. China

[*]Corresponding author:

Yifei Xie: E-mails: xie.yifei2018@gmail.com

Abstract

Species of Elaeocarpaceae vary, but complete chloroplast genome data and systematic comparisons across the family are rarely reported. To understand the variation in chloroplast sequence size and structure in Elaeocarpaceae, the chloroplast genomes of 9 species were sequenced using the Illumina HiSeq 2000 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 5 genera of Elaeocarpaceae. Chloroplast genome characteristics were examined by using Circoletto and IRscope software. The results revealed the following: (a) The 11 sequenced chloroplast genomes ranged in size from 157,546 bp to 159,400 bp. (b) The chloroplast genomes of *Elaeocarpus*, *Sloanea, Crinodendron* and *Vallea* lacked the *rpl32* gene in the small single-copy (SSC) region. The large single-copy (LSC) region of the chloroplast genomes lacked the *ndhK* gene in *Elaeocarpus*, *Vallea stipularis,* and *Aristotelia fruticosa*. The LSC region of the chloroplast genomes lacked the *inf*A gene in *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.

Key words:

Elaeocarpus, Chloroplast, Genome, Divergence time

1. Introduction

In land plants, most chloroplast genomes are single-stranded, circular, or double-stranded DNA sequences. The 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)**. These regions are involved in photosynthesis, transcription, and translation, among other functions **(Gao et al., 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., 2003a; 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 **(Wang and Messing, 2010; Goremykin et al., 2005; Doyle et al., 1995)**. In heterotrophic plants, pseudogenization and entire losses

of *ndh* genes were also detected **(Wickett et al., 2008; Barrett et al., 2014; Wicke et al., 2016)**. However, *ndh* gene loss events have also occurred in autotrophic orchids, gnetophytes and Pinaceae **(Braukmann et al., 2009; Kim et al., 2015; Wakasugi et al., 1994)**. 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, 2016)**. Recent studies suggest that Elaeocarpaceae is a sister group to Cephalotaceae and Brunellia **(Heibl and Renner, 2012; Magallon et al., 2015; Harris and Davies, 2016)**. Moreover, the family can be divided into three monophyletic groups (*Sloanea* alliance, *Tremandraceous* genera and *Elaeocarpus* alliance) according to phylogenetic analysis of multifragment genes **(Phoon, 2015)**. While as the taxonomy of Elaeocarpaceae belonged to Oxalidales was concerned (Savolainen et al., 2000; Soltis et al., 2000; Byng et al., 2016), the relationships within

Oxalidales need a further study for a quite different morphological character (the filaments are longer than the anthers in bud in Oxalidales but not in Elaeocarpaceae) **(Matthews and Endress, 2002)** and limited sampling in certificating the relationships within Oxalidales (Li et al., 2019). Furthermore, the age of genera within Elaeocarpaceae has a visible change based on multigene phylogenies of *trn*L-*trn*F+*trn*V-*ndh*C regions and *trn*L-*trn*F+*ITS* regions **(Crayn et al., 2006; Phoon, 2015)**.

Here we used the whole chloroplast genome sequences to further explore in detail phylogenetic relationships within Elaeocarpaceae and other relative families. This study aims to (1) test genetic category between different genus within Elaeocarpaceae, (2) determine the relationships of Elaeocarpaceae within Oxalidales, (3) 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.

2. Materials and Methods

2.1 Plant material and chloroplast genome sequencing

Leaf materials were sampled from 9 species representing 5 genera of Elaeocarpaceae and collected from a 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 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) 2000 platform in Beijing, China. After removing the low-quality reads from the raw data, we obtained clean data and uploaded them to the NCBI SRA database in fastq format.

2.2 Genome annotation and comparison

The paired-end reads were filtered using the GetOrganelle pipeline (https://github.com/Kinggerm/GetOrganelle) to obtain plastid-like reads, and then the filtered reads were assembled using SPAdes version 3.10

127 **(Bankevich et al., 2012)**. To retain pure chloroplast contigs, the final
128 "fastg" files were filtered using the "slim" script of GetOrganelle. The
129 filtered De Brujin graphs were viewed and edited using Bandage **(Wick**
130 **et al., 2015)**, and then a circular chloroplast genome was generated. We
131 used Plastid Genome Annotator to annotate the published plastid genome
132 of *Elaeocarpus japonicus* (MT985378) as a reference **(Kearse et al.,**
133 **2012)**. The annotated chloroplast genomes have been submitted to
134 GenBank (table 1). Maps of all 9 chloroplast genomes were drawn by the
135 Organellar Genome DRAW tool **(Lohse et al., 2013),** and a map of
136 shared protein-coding genes was drawn by a Venn diagram viewer
137 (http://jvenn.toulouse.inra.fr/app/example.html; **Philippe et al., 2014**).
138 mVISTA online tools (http://genome.lbl.gov/vista/mvista/) were used to
139 determine chloroplast genome similarity among Elaeocarpaceae **(Frazer**
140 **et al., 2004)**. The similarity, rearrangement and inversion of gene blocks
141 were analyzed by Circoletto (http://tools.bat.infspire.org/circoletto/;
142 **Darzentas et al., 2010**). IRscope (https://irscope.shinyapps.io/irapp/) can
143 be useful for assessing IR expansion and contraction in the evolution of
144 chloroplast genomes **(Ali et al. 2018)**.

145

146 **2.3 Phylogenomics and molecular clock dating analysis**

147 To infer phylogenetic relationships within the Elaeocarpaceae and
148 other related families, 20 species of 6 families including Elaeocarpaceae,
149 Cephalotaceae, Brunelliaceae, Oxalidaceae and Connaraceae were
150 compared. The genomes from the 6 families included 11 new chloroplast
151 genomes and 9 published complete chloroplast genome (table 1), that of
152 *Euonymus schensianus* (NC036019) and *Euonymus maackii*
153 (MW771518), which was obtained from the NCBI database and treated
154 as the outgroup **(Baker et al., 2021; Li et al., 2021)**. For the species tree,
155 Bayesian inference (BI) analyses were performed on data sets of 20
156 chloroplast genome sequences. The whole-genome matrix was aligned
157 using MAFFT version 3.73 **(Katoh and Standley, 2013)** and then
158 manually edited using Geneious version 9.1.7 **(Kearse et al. 2012).**
159 Bayesian inference (BI) was performed using MrBayes version 3.2.6
160 **(Ronquist and Huelsenbeck, 2003)**. The best-fitting DNA substitution
161 model according to the Bayesian information criterion (BIC), GTR + F+ I,
162 was identified by using jModelTest version 2.1.10 **(Darriba et al., 2012;**

163 **Guindon et al., 2003)**. Markov chain Monte Carlo (MCMC) analyses
164 were run in MrBayes for 10,000,000 generations. The BI analysis started
165 with a random tree and sampled trees every 1,000 generations. The first
166 25% of the trees were discarded as burn-in, and the remaining trees were
167 used to generate a majority-rule consensus tree. Besieds, we also
168 estimated a maximum likelihood phylogeny for the genera in RAxML
169 v8.0.0 **(Stamatakis et al., 2008)**, on the CIPRES web server
170 (www.phylo.org). We used the default settings, including a TVM + R3 +
171 F model of sequence evolution,
172 Then, based on BEAST 1.10.4, a lognormal relaxed clock model
173 was run by using the GTR + F+ I site model with four gamma categories,
174 with a random starting tree and a Yule speciation process tree prior
175 **(Suchard et al., 2018)**. MCMC was performed with 500 million
176 generations and sampling every 50,000 generation and the effective
177 sample size (ESS) values was confirmed exceeded 200 for all parameters.
178 Then we used the phyutility software to generate an all-compatible
179 consensus tree **(Smith and Dunn, 2008)**. Node ages were optimised onto
180 this consensus phylogeny as the median value for a given node across all

trees in the posterior distribution that contained the node using the TreeAnnotator software **(Drummond and Rambaut, 2007)**. Additionally, the phylogeny was calibrated using 4 fossils, fossil one from a related clade and by setting the split between *Sloanea* and *Vallea* to 55±2 Ma **(Mayr 2000)**. We used the 40±10 Ma split between *Vallea* and *Aristotelia* as the calibration point **(Heibl and Renner, 2012)**. fossil three is *Elaeocarpus* from the Tasmania in Austalia that is about 55±2 Ma old **(Hill 1984)**. The other fossil are leaves of *Rourea* (Connaraceae) from Panama, dated to 59 Ma **(Graham 1988)**.The tree was viewed and edited with FigTree version 1.4.0 software (http://tree.bio.ed.ac.uk/software/figtree/).

Table 1. Sampled species and their voucher specimens from the NCBI database.

3. Results

3.1 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 (*Sloanea sinensis*) to 159,400 bp (*Crinodendron patagua*). The length of the LSC region ranged from 85,874 bp (*Elaeocarpus japonicus*) to 88,413 bp (*Sloanea sinensis*), that of the IR regions ranged from 25,984 bp (*Sloanea sinensis)* to 27,437 bp (*Elaeocarpus japonicus* and *Elaeocarpus japonicus* var. *yunnanensis*), and that of the SSC region ranged from 16,981 bp (*Elaeocarpus japonicus*) to 17,958 bp (*Crinodendron Patagua*). The total GC content of the 11 chloroplast genomes from 5 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 *Aristotelia fruticosa,* which had 133 genes, including 8 rRNA genes, 37 tRNA genes, and 88 protein-coding genes, the chloroplast genomes of the other 4 genera had 132 genes, including 8 rRNA genes, 37 tRNA genes, and 87 protein-coding genes. A total of 114

217 unique genes were detected in the chloroplast genome of *Aristotelia*
218 *fruticosa*, while *Crinodendron patagua*, *Vallea stipularis* and the genus
219 *Sloanea* had 113 unique genes, and genus *Elaeocarpus* had 111 unique
220 genes (table 2).
221
222 **3.2 Chloroplast genome comparisons**
223 The 5 genera shared 111 protein-coding genes, but *rpl*32 was
224 detected in the SSC region of only the chloroplast genome of *Aristotelia*
225 *fruticosa* (fig. 1). The LSC region of the chloroplast genomes lacked
226 *ndhK* in the genus *Elaeocarpus*, *Vallea stipularis*, and *Aristotelia*
227 *fruticosa* and lacked *inf*A in the genus *Elaeocarpus* and *Crinodendron*
228 *patagua*. *ycf*68 was found in *Vallea stipularis*, *Aristotelia fruticosa* and
229 *Crinodendron patagua.* In addition, synteny was detected in the 5 genera
230 of Elaeocarpaceae (fig. 2). A significant degree of synteny was found
231 between *Vallea stipularis* and *Aristotelia fruticosa*, *Elaeocarpus* and
232 *Sloanea*. However, the synteny between *Crinodendron patagua* and the
233 other 4 genera was low. Five genera of Elaeocarpaceae were compared, in
234 addition to the species in *Elaeocarpus* and *Sloanea*. Two groups,

Elaeocarpus angustifolius and *Elaeocarpus hainanensis* as well as *Elaeocarpus japonicus* and *Elaeocarpus japonicus* var. *yunnanensis*, had more blocks of synteny in the genus *Elaeocarpus*. Several blocks of synteny were detected in the 4 chloroplast genomes of the genus *Sloanea*, suggesting that the 4 species are similar to each other.

3.3 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 *Vallea stipularis*, *Aristotelia fruticosa*, *Crinodendron 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 *Elaeocarpus angustifolius* was 370 bp, much greater than that in the chloroplast genomes of other species in the genus *Elaeocarpus*: *Elaeocarpus japonicus* var. *yunnanensis*, *Elaeocarpus japonicus* and *Elaeocarpus angustifolius*.

253 Moreover, the lengths of *ndh*F and *ycf*1 in *Elaeocarpus angustifolius*
254 were shorter than those in the other three species. For the genus *Sloanea*,
255 the chloroplast genomes of 4 species, *Sloanea sinensis*, *Sloanea*
256 *cordifolia*, *Sloanea dasycarpa* and *Sloanea longiaculeatae*, were
257 generally the same in terms of IR expansion and contraction, with the
258 exception that the length of *ycf*1 in *Sloanea dasycarpa* and *Sloanea*
259 *longiaculeatae* was greater than that in *Sloanea sinensis* and *Sloanea*
260 *cordifolia*.
261
262 **3.4 Phylogenomics and molecular clock dating analysis**
263 The matrix of complete chloroplast genomes was used to reconstruct
264 a phylogenomic tree of Oxalidales (fig. 4). The molecular tree showed
265 that *Rourea_microphylla* representing Connaraceae started to diversify at
266 119 Mya. *Averrhoa carambola*, *Oxalis_drummondii* and
267 *Oxalis_corniculata* representing Oxalidaceae diverged from Oxalidales
268 (ca. 122 Mya) at 73 Mya. Cephalotaceae (ca. 60 Mya) has the closest
269 genetic relationship with Elaeocarpaceae (ca. 60 Mya) with that of
270 Oxalidaceae (ca. 119 Mya). Brunelliaceae has a similar differentiation

time about 60 Mya from Oxalidaceae.

The molecular tree also showed the sister relationships of 11 chloroplast genomes from 5 representative genera of Elaeocarpaceae was highly supported. Clade I, containing the genus *Elaeocarpus,* was 100% supported and was dated to ca. 53 Mya, and the crown node age of clade II (*Crinodendron patagua*) was dated to ca. 55 Mya. Diversification of clade III, containing the *Sloanea* alliance (*Vallea stipularis*, *Aristotelia fruticosa* and the genus *Sloanea*), was dated to 53 Mya. Further differentiation of *Vallea stipularis* and *Aristotelia fruticosa* took place within the last 3 Mya. In addition, the genus *Sloanea* started to diversify during the late Miocene (ca. 0.4 Mya).

4. Discussion

4.1 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

289 is, the higher the density of DNA and the more conserved the chloroplast
290 genome **(Do et al., 2013)**. Therefore, variation might occur in the SSC
291 region rather than the IR regions. Comparisons of the 11 plastomes
292 showed the loss of *inf*A in *Crinodendron patagua* and the genus
293 *Elaeocarpus*, and similar losses or pseudogenization was reported in the
294 309 complete chloroplast genomes of 24 species of angiosperms **(Millen**
295 **et al., 2001)**. *ndh* genes are frequently pseudogenized or lost in plant
296 groups with a degree of heterotrophy due to evolutionary adaptation to
297 excessive water in the environment, as observed in *Aneura*, *Cuscuta*,
298 *Epifagus*, *Hydnora*, and nonphotosynthetic orchid species and some
299 autotrophic gymnosperms and ferns **(De Pamphilis and Palmer, 1990;**
300 **Wicke et al., 2011; Wickett et al., 2008; McNeal et al., 2007; Kim et**
301 **al., 2015; Naumann et al., 2016),** and this study also revealed that
302 *Aristotelia fruticosa*, *Vallea stipularis* and the genus *Elaeocarpus* of
303 Elaeocarpaceae have lost the *ndh*K gene. The *rpl*32 gene was detected in
304 *Aristotelia fruticosa* but not in the other 4 genera (*Vallea stipularis*,
305 *Crinodendron patagua*, the genus *Elaeocarpus* and the genus *Sloanea*),
306 which is similar to previously published research about the losses of two

307 genes, *inf*A and *rpl*32, in *Thalictrum coreanum* **(Park and Jansen, 2015)**.
308 In summary, the 5 genera may have experienced different niche
309 expansions.

310 Some studies suggest that IR expansion and contraction are
311 associated with the evolution of plants. Large expansions and
312 contractions may be related to double-strand breakage and repair, while
313 the small expansions and contractions may be related to gene conversions
314 **(Khakhlova et al., 2006; Liang, 2018; Hansen et al., 2007; Kim and**
315 **Lee, 2004; Wang et al., 2008; Ma et al., 2013)**. We found large IR
316 expansions in the 5 genera. The genus *Elaeocarpus* is different from the
317 other 4 genera at the IR/SC boundary, which may reflect that the genus
318 *Sloanea* has an older origin and experienced a different evolution event.
319 In addition, *rps*19 was located across the LSC/IRB regions in 4 genera,
320 while the boundary of the LSC and IRb regions in the genus *Elaeocarpus*
321 included *rps*3. Research shows that the locations of *rps*19 and *rps*3 differ
322 between the chloroplasts of monocotyledons and dicotyledons. In some
323 dicotyledons, *rps*19 only partially exists in the IR region, while the *rps*3
324 gene is only found in *Paris* and Melanthiaceae **(Lin et al., 2012; Sarah**

et al., 2013). Compared with the other 4 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 and 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 *Vallea stipularis* and *Aristotelia fruticosa*, the genus *Elaeocarpus* and the genus *Sloanea,* meaning that the times of divergence between the genus *Sloanea* and the genus *Elaeocarpus*, *Vallea stipularis* and *Aristotelia fruticosa* were similar. Interestingly, there were no blocks of synteny in *Crinodendron patagua* with the other 4 genera, meaning that the evolution of *Crinodendron patagua* was different from that of the other 4 genera. In the genus *Elaeocarpus* and genus *Sloanea*, it is worth noting that the differentiation time of *Elaeocarpus japonicus* was similar to that

of *Elaeocarpus japonicus* var. *yunnanensis* and that of *Elaeocarpus angustifolius* was similar to that of *Elaeocarpus hainanensis*. In addition, the times of divergence among species in the genus *Sloanea* were similar.

4.2 Phylogenomic relationships and historical biogeography in Oxalidales

Based on the 20 species of 6 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)**. The 5 genera of Elaeocarpaceae were clarified as sister to Cephalotaceae and Brunelliaceae; In addition, the family Connaraceae and Oxalidaceae are far from Elaeocarpaceae, which was recognized by **Heibl and Renner (2012)**.

Pillon's (2021) phylogeny of Oxalidales based on DNA Molecular fragments has been used as data for event-based biogeographic analysis of the world. In that study the possibly ancestral area for the Oxalidales is Australia/New Guinea + New Caledonia in Cretaceous (102 Mya), which was consistent with our result (120 Mya). That also can be verified by the

361 greatest number of extant species and genera in Oceania, and particularly
362 in eastern Australia, New Guinea, and New Caledonia **(Kershaw et al.,**
363 **1976; Kershaw et al., 2007; Sniderman, 2011)**.

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

371 The differentiation time of Oxalidaceae is consistent as that of **Heibl**
372 **and Renner (2012)**, which is about 68 Mya. Geographical distribution
373 patterns suggest the origin of the family in the southern hemisphere, prior
374 to the separation of South America and Africa **(Raven and Axelrod,**
375 **1974)**.

376 The split from Cephalotaceae and Brunelliaceae was estimated at 60
377 Mya, more recent than **Heibl and Renner's (2012)** reasearch (78 Mya).
378 Brunellia is exclusively distribute in the continent of America, and most

379 of the species distribute in North America, but with only 6 of the known
380 species (61 species) occurring in north of Panama. The presence of
381 Brunellia may have been represented north of Panama before the closing
382 of the central American land bridge **(Montes et al., 2012,** 15Mya**)**, which
383 was consistent with our result **(Coode, 2004).**

384 It has long been postulated that Elaeocarpaceae originated in the
385 southern hemisphere, of which only *Elaeocarpus* and *Sloanea* reach the
386 northern hemisphere (Raven and Axelrod, 1974). The ages of
387 Elaeocarpaceae (60Mya) estimated in this study were younger than early
388 results estimated at 79.62-85.2 Mya **(Magallón et al., 2015; Phoon,**
389 **2015)**, 64-66 Mya **(Wikström et al., 2001)**, 67 Mya **(Heibl and Renner,**
390 **2012)** and 100 Mya **(Crayn et al., 2006)**, but older than 38 Mya **(Harris**
391 **and Davies, 2016)**. These differences may be due to the choice of DNA
392 markers and the accuracy of the fossil calibrations of molecular
393 evolutionary rates. The earliest divergence within the Elaeocarpaceae
394 appears to have occurred in the late Cretaceous based on our data, which
395 is broadly coincident with the time when the western (Africa and South
396 America) and eastern (Australia, Antarctica, Madagascar, and India) parts

of Gondwana were separating **(Ali and Aitchison, 2008)**.

4.3 Phylogenomic relationships and historical biogeography in Elaeocarpaceae

Within Elaeocarpaceae, the 11 taxa were separated into the following groups in our study: the *Sloanea* alliance (*Vallea stipularis*, *Aristotelia fruticosa* and *Sloanea*), *Elaeocarpus* alliance and *Crinodendron 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 + *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.* The clades identified in the phylogenomic analyses strongly confirmed the preliminary results of earlier studies, and the results of the analysis improved the posterior probabilities of all

clades **(Maynard, 2004; Baba, 2013; Phoon, 2015)**.

Compared with the differentiation of *Vallea stipularis* and *Aristotelia fruticosa* in **Phoon's (2015)** study, the age of the split between *Vallea stipularis* and *Aristotelia fruticosa* was much younger than the age estimated at 37 Mya. The results of the present study agreed with Coode's phylogenetic reconstruction in which *Vallea* and *Aristotelia* were sister group and the ancestors may have dispersed between western and eastern Gondwana **(Coode, 2004; Phoon, 2015)**. The minimum estimates of divergence times between *Vallea stipularis* and *Aristotelia 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 dwarf cloud forest near the Atlantic coast of Brazil) as sister to the rest of

433 the genus, based on morphological data **(Coode, 2004)**.

434 *Elaeocarpus* represents a widespread lineage in Elaeocarpaceae that
435 diverged 53 Mya, which was more older than Phoon's estimate (40 Mya).
436 Divergence time analysis suggests that *Elaeocarpus* split in the Eocene
437 and migrated out of Australia to the surrounding regions mostly in the
438 Oligocene and the Miocene should be doubted as sampling without
439 species from Southeast Asian **(Crayn et al., 2006; Heibl and Renner,**
440 **2012; Phoon, 2015)**.

441 Divergence time analysis using BEAST suggests that *Sloanea*
442 diverged from its sister species *Vallea stipularis* and *Aristotelia fruticosa*
443 at 0.4 Mya, more recent than 29 Mya **(Phoon, 2015)**, obviously a
444 deviation caused by the sample represents the *Sloanea* in East Asia.

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

449

450

Table 2. Summary of 11 complete chloroplast genomes of Elaeocarpaceae

Acknowledgments

The authors would like to acknowledge Dr. Hai-Zhong Yu and Prof. Ren-Lin Liu from Gannan Normal University. This work was supported by the National Natural Science Foundation of China (grant numbers 31110103911, J1310002), Science and technology project in Jiangxi province department of education (GJJ170839, 190781).

Conflict Of Interest

The authors declare no conflicts of interest regarding publication of this paper.

CRediT authorship contribution statement

Yihui Wang: Formal analysis (Lead), Data curation (Lead), Writing - original draft(Lead), Methodology (Lead), Resources (Equal), Software (Lead), Visualization (Lead).

Yifei Xie: Conceptualization (Lead), Project administration (Lead), Writing - review & editing (Lead), Funding acquisition (Lead).

Jiayi Jin: Methodology (Equal).

Jinyue Li: Software (Equal).

Xiangdong Qiu: Resources (Supporting).

Yang Tong: Resources (Supporting).

Zhongyang Li: Funding acquisition (Equal), Resources (Equal).

Zhixiang Zhang: Resources (Supporting), Funding acquisition (Equal).

Wenling Lai: Resources (Supporting), Funding acquisition (Supporting).

Data Availability Statement

The original sequencing data have been submitted to the NCBI data base and received GenBank accession numbers MT982368 (*Aristotelia fruticosa*), MT982369 (*Crinodendron patagua*), MT982370 (*Vallea stipularis*), MW242787 (*Elaeocarpus angustifolius*), MW602804

(*Elaeocarpus hainanensis*), MT683335 (*Elaeocarpus japonicus*), MW242788 (*Elaeocarpus japonicus* var. *yunnanensis*), MW004670 (*Sloanea sinensis*), MW242789 (*Sloanea cordifolia*), MW242790 (*Sloanea dasycarpa*), MW242791 (*Sloanea longiaculeatae*). The data used to support the findings of this study are included in Table 1.

Fig. 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*,

505 *Aristotelia fruticosa* and *Vallea stipularis.* The brown block represents the
506 genes unique to *Aristotelia fruticosa*, *Crinodendron patagua* and *Vallea*
507 *stipularis*.)
508
509 **Fig. 2** - Synteny detected in Elaeocarpaceae using Circoletto.
510 (**A** - Synteny detected between the chloroplast genomes of the
511 Elaeocarpaceae species *Aristotelia fruticosa*, *Crinodendron patagua*,
512 *Vallea stipularis*, *Elaeocarpus japonicus* and *Sloanea sinensis* using
513 Circoletto. **B** - Synteny detected between the chloroplast genomes of the
514 Elaeocarpaceae species *Elaeocarpus angustifolius, Elaeocarpus*
515 *japonicus, Elaeocarpus japonicus* var. *yunnanensis* and *Elaeocarpus*
516 *hainanensis* using Circoletto. **C** - Synteny detected between the
517 chloroplast genomes of the Elaeocarpaceae species *Sloanea cordifolia,*
518 *Sloanea dasycarpa, Sloanea longiaculeatae* and *Sloanea sinensis* using
519 Circoletto.)
520
521 **Fig. 3** - Comparisons of IR expansion and contraction in Elaeocarpaceae.
522 (**A** - The chloroplast genome boundaries of *Aristotelia fruticosa,*

523 *Crinodendron patagua, Vallea stipularis, Elaeocarpus japonicus* and
524 *Sloanea sinensis* of Elaeocarpaceae. **B** - The chloroplast genome
525 boundaries of *Elaeocarpus angustifolius*, *Elaeocarpus japonicus,*
526 *Elaeocarpus japonicus* var. *yunnanensis* and *Elaeocarpus hainanensis.* **C**
527 - The chloroplast genome boundaries of *Sloanea cordifolia, Sloanea*
528 *dasycarpa, Sloanea longiaculeatae* and *Sloanea sinensis.*)
529
530 **Fig. 4** - Molecular phylogenomic tree of 20 species of Oxalidales.
531 (Molecular phylogenomic tree of 20 species of Oxalidales based on
532 complete chloroplast genome sequences constructed using Bayesian
533 inference (BI) and Maximum Likelihood (ML). Numbers at each node
534 are bootstrap support values and posterior probability.)
535
536
537
538
539
540

Literature Cited

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

Ali, J.R., Aitchison, J.C., 2008. Gondwana to Asia: plate tec-tonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through end Eocene (166-35 Ma). Earth-Science Reviews. 88, 145-166. https://doi.org/10.1016/j.earscirev.2008.01.007.

Alzahrani, D.A., Enas, J.A., Samaila, S.Y., Abba, A., 2021. Comparative Analysis of Chloroplast Genomes of Four Medicinal Capparaceae Species: Genome Structures, Phylogenetic Relationships and Adaptive Evolution. Plants. 10(6), 1229-1229. https://doi.org/10.3390/plants10061229.

Baas, P., Manchester, S., Wheeler, E. A., Srivastava, R., 2017. Fossil wood with dimorphic fibres from the Deccan Intertrappean Beds of India-the oldest fossil Connaraceae? IAWA Journal. 38, 124-133.

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

Baker, W. J., Bailey, P., Barber, V., Barker, A., Bellot, S., Bishop, D., Botigué, L. R., Brewer, G., Carruthers, T., Clarkson, J. J., Cook, J., Cowan, R. S., Dodsworth, S., Epitawalage, N., Françoso, E., Gallego, B., Johnson, M. G., Kim, J. T., Leempoel, K., Maurin, O., McGinnie, C., Pokorny, L., Roy, S., Stone, M., Toledo, E., Wickett, N. J., Zuntini, A. R., Eiserhardt ,W. L., Kersey, P. J., Leitch, I. J., Forest, F., 2021. A Comprehensive Phylogenomic Platform for Exploring the Angiosperm Tree of Life. Syst Biol. syab035. https://doi.org/10.1093/sysbio/syab035.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-ce ll sequencing. J. Comput. Biol. 19(5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

575 Barrett, C.F., Freudenstein, J.V., Li, J., Mayfield-Jones, D.R., Perez, L.,
576 Pires, J.C., Santos, C., 2014. Investigating the path of plastid
577 genome degradation in an early-transitional clade of heterotrophic
578 orchids, and implications for heterotrophic angiosperms. Mol. Biol.
579 Evol. 31(12), 3095–3112. https://doi.org/10.1093/molbev/msu252.

580 Bock, R., 2007. Plastid biotechnology: prospects for herbicide and insect
581 resistance, metabolic engineering, and molecular farming. Curr.
582 Opin. Biotechnol. 18(2), 100–106.
583 https://doi.org/10.1016/j.copbio.2006.12.001.

584 Braukmann, T.W.A., Kuzmina, M., Stefanovic, S., 2009. Loss of all
585 plastid *ndh* genes in Gnetales and conifers: extent and evolutionary
586 significance for the seed plant phylogeny. Curr. Genet. 55(3),
587 323–337. https://doi.org/10.1007/s00294-009-0249-7.

588 Byng, J.W., Chase, M.W., Maarten, J.M., Christenhusz, M.J.M., Fay, M.
589 F., Judd, W.S., Soltis, D.E., Mabberley, D.J.A., Sennikov, N., Soltis,
590 P.S., Stevens, P.F., Barbara, B., Brockington, S., Chautems, A.,
591 Clark, J.C., Conran, J., Haston, E., Möller, M., Moore, M.,

592 Olmstead, R., Perret M., Skog, L., Smith, J., Tank, D., Vorontsova,
593 M., Weber, A., 2016. An update of the Angiosperm Phylogeny
594 Group classification for the orders and families of flowering plants:
595 APG IV. Bot. J. Linn. Soc. 181(1), 1–20.
596 https://doi.org/10.1111/boj.12385.

597 Cai, Z., Penaflor, C., Kuehl, J.V., Leebens-Mack, J., Carlson J.E.,
598 dePamphilis, C.W., Boore, J.L., Jansen, R.K., 2006. Complete
599 plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*:
600 implications for the phylogenetic relationships of magnoliids. BMC
601 Evol. Biol. 6, 77. https://doi.org/10.1186/1471-2148-6-77.

602 Cauz-Santos, L.A., Da-Costa. Z.P., Callot C., Cauet S., Zucchi, M. I.,
603 Bergès, H., Berg, C.V.D., Lucia, M., Vieira, C., 2020. A repertory of
604 rearrangements and the loss of an inverted repeat region in Passiflora
605 chloroplast genomes. Genome. Biol. Evol. 12(10), 1841-1857.
606 https://doi.org/10.1093/gbe/evaa155.

607 Cheng, F., Mandakova, T., Wu, J., Xie, Q., Lysak, M.A., Wang, X., 2013.
608 Deciphering the diploid ancestral genome of the mesohexaploid

609 *Brassica rapa*. The Plant Cell Online. 25(5), 1541-1554.
610 https://doi.org/10.1105/tpc.113.110486.

611 Christenhusz, M., Byng, J.W., 2016. The number of known plant species
612 in the world and its annual increase. Phytotaxa. 261, 201-217.
613 https://doi.org/10.11646/phytotaxa.261.3.1.

614 Coode, M.J.E., 2004. Families and Genera of Vascular Plants, in:
615 Kubitzki, K., (Eds.), Elaeocarpaceae, pp. 135-143.

616 Crayn, D.M., Rossetto, M., Maynard, D.J., 2006. Molecular phylogeny
617 and dating reveal an Oligo-Miocene radiation of dry-adapted shrubs
618 (Tremandraceae) from rainforest tree progenitors (Elaeocarpaceae)
619 in Australia. Amer. J. Bot. 93 (9), 1328-1342.
620 https://doi.org/10.3732/ajb.93.9.1328.

621 Gao, L., Su, Y.J., Wang, T., 2010. Plastid genome sequencing,
622 comparative genomics, and phylogenomics: Current status and
623 prospects. J. Syst. Evol. 48(2), 77–93.
624 https://doi.org/10.1111/j.1759-6831.2010.00071.x.

625 Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2:
626 more models, new heuristics, and parallel computing. Nat. Methods.
627 9(8), 772-772. https://doi.org/10.1038/nmeth.2109.

628 Darzentas, N., 2010. Circoletto: visualizing sequence similarity with
629 Circos. Bioinformatics. 26(20), 2620-1.
630 https://doi.org/10.1093/bioinformatics/btq484.

631 DePamphilis, C.W., Palmer, J.D., 1990. Loss of photosynthetic and
632 chlororespiratory genes from the plastid genome of a parasitic
633 flowering plant. Nature. 348(6299), 337–339.
634 https://doi.org/10.1038/348337a0.

635 Do, H.D.K., Kim, J.S., Kim, J.H., 2013. Comparative genomics of four
636 Liliales families inferred from the complete chloroplast genome
637 sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). Gene.
638 530(2), 229-235. https://doi.org/10.1016/j.gene.2013.07.100.

639 Doyle, J.J., Doyle, J.L., Palmer, J.D., 1995. Multiple independent losses
640 of 2 genes and one intron from legume chloroplast genomes. Syst.
641 Bot. 20(3), 272–294. https://doi.org/10.2307/2419496.

642 Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary
643 analysis by sampling trees. BMC Evol Biol. 8(7), 214.
644 https://10.1186/1471-2148-7-214.

645 Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004.
646 VISTA: computational tools for comparative genomics. Nucleic
647 Acids Res. 32 (Web Server issue), W273-9.
648 https://doi.org/10.1093/nar/gkh458.

649 Goremykin, V., Hirsch-Ernst, K., Wölfl, S., Hellwig, F.H., 2003a. The
650 chloroplast genome of the “basal” angiosperm *Calycanthus fertilis* –
651 structural and phylogenetic analyses. Plant. Syst. Evol. 242(1),
652 119–135. https://doi.org/10.1007/s00606-003-0056-4.

653 Goremykin, V.V., Holland, B., Hirsch-Ernst, K.I., Hellwig, F.H., 2005.
654 Analysis of *Acorus calamus* chloroplast genome and its phylogenetic
655 implications. Mol. Biol. Evol. 22(9), 1813–1822.
656 https://doi.org/10.1093/molbev/msi173.

657 Graham, A., 1988. Studies in Neotropical Paleobotany. V. The Lower
658 Miocene Communities of Panama-The Culebra Formation. Annals

659 of the Missouri Botanical Garden. 75(4), 1440-1466.
660 https://doi.org/10.2307/2399295.

661 Guindon, S., Gascuel, O., Rannala, B., 2003. A simple, fast, and accurate
662 algorithm to estimate large phylogenies by maximum likelihood.
663 Syst. Biol. 52(5), 696–704.
664 https://doi.org/10.1080/10635150390235520.

665 Guisinger, M.M., Chumley, T.W., Kuehl, J.V., Boore, J.L., Jansen, R.K.,
666 2010. Implications of the plastid genome sequence of *Typha*
667 (Typhaceae, Poales) for understanding genome evolution in Poaceae.
668 J. Mol. Evol. 70(2), 149–166.
669 https://doi.org/10.1007/s00239-009-9317-3.

670 Hansen, D.R., Dastidar, S.G., Cai, Z., Penaflor, C., Kuehl, J.V., Boore,
671 J.L., Jansen, R.K., 2007. Phylogenetic and evolutionary implications
672 of complete chloroplast genome sequences of four early-diverging
673 angiosperms: *Buxus* (Buxaceae), *Chloranthus* (Chloranthaceae),
674 *Dioscorea* (Dioscoreaceae), and *Illicium* (Schisandraceae). Mol.
675 Phylogenet. Evol. 45(2), 547-563.

https://doi.org/10.1016/j.ympev.2007.06.004.

Harris, L.W., Davies, T.J., 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. https://doi.org/10.1371/journal.pone.0162907.

Heibl, C., Renner, S.S., 2012. Distribution models and a date phylogeny for chilean *Oxalis* species reveal occupation of new habits by different lineages, not rapid adaptive radiation. Syst. Biol. 61, 823-834. https://doi.org/10.1093/sysbio/sys034.

Hill, R.S., 1984. Tertiary Nothofagus macrofossils from Cethana, Tasmania. Alcheringa: An Australasian Journal of Palaeontology. 8(1), 81-86. https://doi.org/10.1080/03115518408619610.

Karol, K.G., Arumuganathan, K., Boore, J.L., Duffy, A.M., Everett, K.D., Hall, J.D., Hansen, S.K., Kuehl, J.V., Mandoli, D.F., Mishler, B.D., Olmstead, R.G., Renzaglia, K.S., Wolf, P.G., 2010. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early

land plant lineages. BMC Evol. Biol. 10(1), 321. https://doi.org/10.1186/1471-2148-10-321.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol. Evol. 30(4), 772-780. https://doi.org/10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kershaw, A. P., Bretherton, S. C. L., van der Kaars, W. A., 2007. A complete pollen record of the last 230 ka from Lynch's Crater, north-eastern Australia. Palaeogeography, Palaeoclimatology, Palaeoecology. 251(1), 23-45. https://doi.org/10.1016/j.palaeo.2007.02.015

Khakhlova, O., Bock, R., 2006. Elimination of deleterious mutations in

plastid genomes by gene conversion. Plant J. 46(1), 85-94. https://doi.org/10.1111/j.1365-313X.2006.02673.x.

Kim, H.T., Kim, J.S., Moore, M.J., Neubig, K.M., Williams, N.H., Whitten, W.M., Kim, J.H., 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. https://doi.org/10.1371/journal.pone.0142215.

Kim, K.J., Lee, H.L., 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 11(4), 247-261. https://doi.org/10.1093/dnares/11.4.247.

Li, H.T., Yi, T.S., Gao, L.M., Ma, P.F., Zhang, T., Yang, J.B., Gitzendanner, M.A., Fritsch, P.W., Cai, J., Luo, Y., Wang, H., van der Bank, M., Zhang, S.D., Wang, Q.F., Wang, J., Zhang, Z.R., Fu, C.N., Yang, J., Hollingsworth, P.M., Chase, M.W., Soltis, D.E., Soltis, P.S., Li, D.Z., 2019. Origin of angiosperms and the puzzle of

727 the Jurassic gap. Nat. Plants 5. 461-470.
728 https://doi.org/10.1038/s41477-019-0421-0.

729 Li, H.T., Luo, Y., Gan, L., Ma, P.F., Gao, L.M., Yang, J.B., Cai, J.,
730 Gitzendanner, M.A., Fritsch, P.W., Zhang, T., Jin, J.J., Zeng, C.X.,
731 Wang, H., Yu, W.B, Zhang, R., van der Bank, M., Olmstead, R.G.,
732 Hollingsworth, P.M., Chase, M.W., Soltis, D.E., Soltis, P.S., Yi, T.S.,
733 Li, D.Z., 2021. Plastid phylogenomic insights into relationships of
734 all flowering plant families. BMC Biol. 19(1):232. https://doi.org/
735 10.1186/s12915-021-01166-2.

736 Lin, C.P., Wu, C.S., Huang, Y.Y., Chaw S.M., 2012. The complete
737 chloroplast genome of *Ginkgo biloba* reveals the mechanism of
738 inverted repeat contraction. Genome Bio. Evol. 4(3), 374-381.
739 https://doi.org/10.1093/gbe/evs102.

740 Lohse, M., Drechsel, O., Kahlau, S., Bock, R., 2013. Organellar Genome
741 DRAW–a suite of tools for generating physical maps of plastid and
742 mitochondrial genomes and visualizing expression data sets. Nucleic.
743 Acids Res. 41(Web Server issue), W575–W581.

744 https://doi.org/10.1093/nar/gkt289.

745 Magallon, S., Gómez, A.S., Sánchez, R., Luna, L., Hernández, T., 2015. A
746 metacalibrated time-tree documents the early rise of flowering plant
747 phylogenetic diversity. New phytologist. 207(2), 437-453.
748 https://doi.org/10.1111/nph.13264.

749 Matthews, M. L., Endress, P. K., 2002. Comparative structure and
750 systematics in Oxalidales (Oxalidaceae, Connaraceae, Brunelliaceae,
751 Cephalotaceae, Cunoniaceae, Elaeocarpaceae, Tremandraceae). Bot.
752 J. Linn. Soc. 140, 321-381.

753 Ma, J., Yang, B., Zhu, W., Sun, L., Tian, J., Wang, X., 2013. The complete
754 chloroplast genome sequence of *Mahonia bealei* (Berberidaceae)
755 reveals a significant expansion of the inverted repeat and
756 phylogenetic relationship with other angiosperms. Gene, 528(2),
757 120-131. https://doi.org/10.1016/j.gene.2013.07.037.

758 Maynard, D.J., 2004. A molecular phylogeny for Australian *Elaeocarpus*
759 (Elaeocarpaceae) and the affinities of a putative new taxon. PhD.
760 University of New South Wales, Sydney.

761 Mayr, G., 2000. A new mousebird (Coliiformes: Coliidae) from the
762 Oligocene of Germany. J Ornithol, 141, 85-92.
763 https://doi.org/10.1007/BF01651775.

764 McCoy, S.R., Kuehl, J.V., Boore, J.L., Raubeson, L.A., 2008. The
765 complete plastid genome sequence of *Welwitschia mirabilis*: an
766 unusually compact plastome with accelerated divergence rates. BMC
767 Evol. Biol. 8, 130. https://doi.org/10.1186/1471-2148-8-130.

768 McLoughlin, S., 2001. The breakup history of Gondwana and its impact
769 on pre-Cenozoic floristic provincialism. Australian Journal of
770 Botany, 49: 271-300. https://doi.org/10.1071/BT00023.

771 McNeal, J.R., Kuehl, J.V., Boore, J.L., DePamphilis, C.W., 2007.
772 Complete plastid genome sequences suggest strong selection for
773 retention of photosynthetic genes in the parasitic plant genus
774 *Cuscuta*. BMC Plant Biol. 7, 57.
775 https://doi.org/10.1186/1471-2229-7-57.

776 Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Nga, T., Heggie,
777 L.L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C., Morden, C.W.,

778 Calie, P. J., Jermiin, L.S., Wolfe, K.H., 2001. Many parallel losses of
779 *infA* from chloroplast angiosperm evolution with multiple
780 independent transfers to the nucleus. Plant Cell. 13(3), 645-658.
781 https://doi.org/10.1105/tpc.13.3.645.

782 Montes, C., Bayona, G., Cardona, A., Buchs, D.M., Silva, C.A., Morón,
783 S., Hoyos, N., Ramírez, D.A., Jaramillo, C.A., Valencia, V., 2012.
784 Arc-continent collision and orocline formation: Closing of the
785 Central American seaway, J. Geophys. Res. 117, B04105,
786 doi:10.1029/2011JB008959.

787 Naumann, J., Der, J.P., Wafula, E.K., Jones, S.S.,Wagner, S.T., Honaas,
788 L.A., Ralph, P.E., Bolin, J.F., Maass, E., Neinhuis, C., Wanke, S.,
789 DePamphilis, C.W., 2016. Detecting and characterizing the highly
790 divergent plastid genome of the nonphotosynthetic parasitic plant
791 *Hydnora visseri* (Hydnoraceae). Genome Biol. Evol. 8(2), 345–363.
792 https://doi.org/10.1093/gbe/evv256.

793 Philippe, B., Jérôme, M., Frédéric, E., Djemiel, C., Klopp, C., 2014.
794 jvenn: an interactive Venn diagram viewer. BMC Bioinformatics. 15,

293. https://doi.org/10.1186/1471-2105-15-293.

Park, K.T., Park, S.J., 2021. Phylogenomic Analyses of Hepatica Species and Comparative Analyses Within *Tribe Anemoneae* (Ranunculaceae). Front. Plant Sci. 12, 728. https://doi.org/10.3389/fpls.2021.638580.

Park, S., Jansen, R.K., 2015. Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *rpl32* gene to the nucleus in the ancestor of the subfamily Thalictroideae. BMC Plant Biol. 15, 40. https://doi.org/10.1186/s12870-015-0432-6.

Phoon, S.N., 2015. Systematics and biogeography of *Elaeocarpus* (Elaeocarpaceae). PhD. James Cook University.

Pillon, Y., Hopkins, H. C. F., Maurin, O., Epitawalage, N., Bradford, J., Rogers, Z. S., Baker, W. J., and Forest, F., 2021. Phylogenomics and biogeography of Cunoniaceae (Oxalidales) with complete generic sampling and taxonomic realignments. American Journal of Botany. 108(7): 1181-1200. https://doi.org/10.1002/ajb2.1688.

811 Raven, P., Axelrod, D.I., 1974. Angiosperm biogeography and past
812 continental movements. Ann. Missouri Bot. Gard. 61:539-673.
813 https://doi.org/10.2307/2395021.

814 Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic
815 inference under mixed models. Bioinformatics 19(12), 1572–1574.
816 https://doi.org/10.1093/bioinformatics/btg180.

817 Roper, J.M., Hansen, S.K., Wolf, P.G., Karol, K.G., 2007. The complete
818 plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm.
819 (Marattiaceae). Amer. Fern J. 97(2), 95–106.
820 https://doi.org/10.1640/0002-8444(2007)97[95: TCPGSO]2.0.CO;2.

821 Sarah, S.S., Kim, J.H., Kim, J.S., 2013. Complete chloroplast genome of
822 *Chionographis japonica* (Willd.) Maxim. (Melanthiaceae):
823 comparative genomics and evaluation of universal primers for
824 Liliales. Pl. Mol. Biol. Rep. 31(6), 1407-1421.
825 https://doi.org/10.1007/s11105-013-0616-x.

826 Savolainen, V., Fay, M.F., Albach, D.C., Backlund, A., van der Bank, M.,
827 Cameron, K.M., Johnson, S.A., Lledó, M.D., Pintaud, J.C., Powell,

M., Sheanhan, M.C., Soltis, D.E., Soltis, P.S., Weston, P., Whitten, W.M., Wurdack, K.J., Chase, M.W., 2000. Phylogeny of the eudicots: a nearly complete familial analysis based on *rbc*L gene sequences. Kew Bulletin. 55, 257-309. https://doi.org/10.2307/4115644.

Smith, S.A., Dunn, C.W., 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. Bioinformatics. 24, 715-716. https://10.1093/bioinformatics/btm619.

Sniderman, J.M.K., 2011. Early Pleistocene vegetation change in upland south-eastern Australia. Journal of Biogeography. 38, 1456-1470. https://doi.org/10.1111/j.1365-2699.2011.02518.x

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J., Nixon, K.C., Farris, J.S., 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbc*L, and *atp*B sequences. Botanical Journal of the Linnean Society. 133, 381-461. https://doi.org/10.1111/j.1095-8339.2000.tb01588.x.

Stamatakis, A., Hoover, P., Rougemont, J. A., 2008. Rapid bootstrap

845 algorithm for the RAxML web servers. Syst. Biol. 57, 758-771.
846 https://doi.org/10.1080/10635150802429642.

847 Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J.,
848 Rambaut, A., 2018. Bayesian phylogenetic and phylodynamic data
849 integration using BEAST 1.10. Virus Evol. 4, vey016.
850 https://doi.org/10.1093/ve/vey016.

851 Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., Sugiura,
852 M., 1994. Loss of all *ndh* genes as determined by sequencing the
853 entire chloroplast genome of the black pine *Pinus thunbergii*. Proc.
854 Natl. Acad. Sci. U S A. 91(21), 9794–9798.
855 https://doi.org/10.1073/pnas.91.21.9794.

856 Wang, M., Cui, L., Feng, K., Deng, P.C., Du, X.H., Wan, F.H., Weining,
857 S., Nie, X.J., 2018. Analysis of chloroplast genomes features of
858 Asteraceae species. Genom. Appl. Biol, 37(12), 5437-5447.
859 https://doi.org/10.1007/s11105-015-0853-2.

860 Wang, R.J., Cheng, C.L., Chang, C.C., Wu, C.L., Su, T.M., Chaw, S.M.,
861 2008. Dynamics and evolution of the inverted repeat-large single

862 copy junctions in the chloroplast genomes of monocots. BMC Evol.
863 Biol, 8(1), 36-49. https://doi.org/10.1186/1471-2148-8-36.

864 Wang, W., Messing, J., 2011. High-throughput sequencing of three
865 Lemnoideae (duckweeds) chloroplast genomes from total DNA.
866 PLoS ONE. 6(9), e24670.
867 https://doi.org/10.1371/journal.pone.0024670.

868 Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E., 2015. Bandage:
869 interactive visualization of de novo genome assemblies.
870 Bioinformatics 31(20), 3350–3352.
871 https://doi.org/10.1093/bioinformatics/btv383.

872 Wicke, S., Schneeweiss, G.M., DePamphilis, C.W., Müller, K.F., Quandt,
873 D., 2011. The evolution of the plastid chromosome in land plants:
874 gene content, gene order, gene function. Plant Mol. Biol. 76(3–5),
875 273–297. https://doi.org/10.1007/s11103-011-9762-4.

876 Wicke, S., Müller, K.F., DePamphilis, C.W., Quandt, D., 2016.
877 Mechanistic model of evolutionary rate variation en route to a
878 nonphotosynthetic lifestyle in plants. Proc. Nat. Acad. Sci. U S A.

879 113(32), 9045–9050. https://doi.org/10.1073/pnas.1607576113.

880 Wickett, N.J., Zhang, Y., Hansen, S.K., Roper, J.M., Kuehl, J.V., Plock,
881 S.A., Wolf, P.G., DePamphilis, C.W., Boore, J.L., Goffinet, B., 2008.
882 Functional gene losses occur with minimal size reduction in the
883 plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol. Biol.
884 Evol. 25(2), 393–401. https://doi.org/10.1093/molbev/msm267.

885 Wikström, N., Savolainen, V., Chase, M.W., 2001. Evolution of the
886 angiosperms: calibrating the family tree. Proc. Biol. Sci. 268(1482),
887 2211-20. https://doi.org/10.1098/rspb.2001.1782.

888 Wu, C.S., Lai, Y.T., Lin, C.P., Wang, W.N., Chaw, S.M., 2009. Evolution
889 of reduced and compact chloroplast genomes (cpDNAs) in
890 gnetophytes: selection toward a lower-cost strategy. Mol. Phylogenet.
891 Evol. 52(1), 115–124. https://doi.org/10.1016/j.ympev.2008.12.026.

892 Wu, M.L., 2016. Phylogeny and chloroplast genome evolution of
893 Bambusoideae. PhD. Fujian Agriculture and Forestry University.

Table 1. Sampled species and their voucher specimens from the NCBI database

No.	Species	Herbarium Code	Voucher	Location	Collector	Accession Number in GenBank
1	**Aristotelia fruticosa* Hook.f.	K	781	The Royal Botanic Gardens	Stevens	MT982368
2	**Crinodendron patagua* Molina	K	652	The Royal Botanic Gardens	Coode	MT982369
3	**Vallea stipularis* L.f.	K	654	The Royal Botanic Gardens	Merello	MT982370
4	**Elaeocarpus angustifolius* Blume	BJFC	140942	Guangxi Academy of Forestry	Xie YF	MW242787
5	**Elaeocarpus hainanensis* Oliver	GNNU	PVHJX014291	Diaoluo Mountain, Hainan	Wang YH	MW602804
6	**Elaeocarpus japonicus* Sieb. et Zucc.	BJFC	160730004	Wugong Mountain, Jiangxi	Xie YF	MT683335
7	**Elaeocarpus japonicus* var. *yunnanensis* C. Chen & Y. Tang	BJFC	XW1746	Wenshan, Yunnan	Xie YF	MW242788
8	**Sloanea sinensis* (Hance) Hemsl.	BJFC	XW1956	Wenshan, Yunnan	Xie YF	MW004670
9	**Sloanea cordifolia* K. M. Feng ex H. T. Chang	BJFC	XW1958	Wenshan, Yunnan	Xie YF	MW242789
10	**Sloanea dasycarpa* (Benth.) Hemsl.	BJFC	XZ581	Wenshan, Yunnan	Xie YF	MW242790
11	**Sloanea longiaculeatae* Y. F. Xie & Z. X. Zhang	BJFC	XW1986	Wenshan, Yunnan	Xie YF	MW242791
12	*Cephalotus follicularis* Labill.			CZ Plants Nursery		NC042597
13	*Brunellia trianae* Cuatrec.	COL	4015	Cerro del Padre Amaya, Colombia		MN585217
14	*Brunellia antioquensis* (Cuatrec.) Cuatrec.	COL	4001	Cerro del Padre Amaya, Colombia		MN615725
15	*Oxalis corniculata* L.	HUTB		Hainan University		NC051971
16	*Oxalis drummondii* A. Gray		TEX-DJPG722			NC043802

17	*Averrhoa carambola* L.	KUS	2014-0241	Thailand		KX364202
18	*Rourea microphylla* (Hook. & Arn.) Planch.	FJFC	FAFU201909(Li)	Zhangzhou, Fujian	Li XP	MT537171
19	*Euonymus schensianus* Maxim.	WUK	ZXZ16005	Shaanxi		NC036019
20	*Euonymus maackii* Rupr.					MW771518

*: Newly published species sequences

903 **Table 2. 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

904