

## Oberlin Digital Commons at Oberlin

---

Faculty & Staff Scholarship

---

4-5-2013

# Complete Plastid Genome Sequencing of Trochodendraceae Reveals a Significant Expansion of the Inverted Repeat and Suggests a Paleogene Divergence between the Two Extant Species

Yanxia Sun

Michael J. Moore

*Oberlin College*, [Michael.Moore@oberlin.edu](mailto:Michael.Moore@oberlin.edu)

Aiping Meng

Pamela S. Soltis

Douglas E. Soltis

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.oberlin.edu/faculty\\_schol](https://digitalcommons.oberlin.edu/faculty_schol)

---

### Repository Citation

Sun Y-x, Moore MJ, Meng A-p, Soltis PS, Soltis DE, Li, J-q, and Wang, H-c. (2013) Complete Plastid Genome Sequencing of Trochodendraceae Reveals a Significant Expansion of the Inverted Repeat and Suggests a Paleogene Divergence between the Two Extant Species. PLoS ONE 8(4): e60429. doi:10.1371/journal.pone.0060429

This Article is brought to you for free and open access by Digital Commons at Oberlin. It has been accepted for inclusion in Faculty & Staff Scholarship by an authorized administrator of Digital Commons at Oberlin. For more information, please contact [megan.mitchell@oberlin.edu](mailto:megan.mitchell@oberlin.edu).

---

**Authors**

Yanxia Sun, Michael J. Moore, Aiping Meng, Pamela S. Soltis, Douglas E. Soltis, Jianqiang Li, and Hengchang Wang

# Complete Plastid Genome Sequencing of Trochodendraceae Reveals a Significant Expansion of the Inverted Repeat and Suggests a Paleogene Divergence between the Two Extant Species

Yan-xia Sun<sup>1,2</sup>, Michael J. Moore<sup>3</sup>, Ai-ping Meng<sup>1</sup>, Pamela S. Soltis<sup>4</sup>, Douglas E. Soltis<sup>4,5</sup>, Jian-qiang Li<sup>1\*</sup>, Heng-chang Wang<sup>1\*</sup>

**1** Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei, China, **2** University of Chinese Academy of Sciences, Beijing, China, **3** Department of Biology, Oberlin College, Oberlin, Ohio, United States of America, **4** Florida Museum of Natural History, University of Florida, Gainesville, Florida, United States of America, **5** Department of Biology, University of Florida, Gainesville, Florida, United States of America

## Abstract

The early-diverging eudicot order Trochodendrales contains only two monospecific genera, *Tetracentron* and *Trochodendron*. Although an extensive fossil record indicates that the clade is perhaps 100 million years old and was widespread throughout the Northern Hemisphere during the Paleogene and Neogene, the two extant genera are both narrowly distributed in eastern Asia. Recent phylogenetic analyses strongly support a clade of Trochodendrales, Buxales, and *Gunneridae* (core eudicots), but complete plastome analyses do not resolve the relationships among these groups with strong support. However, plastid phylogenomic analyses have not included data for *Tetracentron*. To better resolve basal eudicot relationships and to clarify when the two extant genera of Trochodendrales diverged, we sequenced the complete plastid genome of *Tetracentron sinense* using Illumina technology. The *Tetracentron* and *Trochodendron* plastomes possess the typical gene content and arrangement that characterize most angiosperm plastid genomes, but both genomes have the same unusual ~4 kb expansion of the inverted repeat region to include five genes (*rpl22*, *rps3*, *rpl16*, *rpl14*, and *rps8*) that are normally found in the large single-copy region. Maximum likelihood analyses of an 83-gene, 88 taxon angiosperm data set yield an identical tree topology as previous plastid-based trees, and moderately support the sister relationship between Buxaceae and Gunneridae. Molecular dating analyses suggest that *Tetracentron* and *Trochodendron* diverged between 44–30 million years ago, which is congruent with the fossil record of Trochodendrales and with previous estimates of the divergence time of these two taxa. We also characterize 154 simple sequence repeat loci from the *Tetracentron sinense* and *Trochodendron aralioides* plastomes that will be useful in future studies of population genetic structure for these relict species, both of which are of conservation concern.

**Citation:** Sun Y-x, Moore MJ, Meng A-p, Soltis PS, Soltis DE, et al. (2013) Complete Plastid Genome Sequencing of Trochodendraceae Reveals a Significant Expansion of the Inverted Repeat and Suggests a Paleogene Divergence between the Two Extant Species. PLoS ONE 8(4): e60429. doi:10.1371/journal.pone.0060429

**Editor:** Jonathan H. Badger, J. Craig Venter Institute, United States of America

**Received:** December 15, 2012; **Accepted:** February 26, 2013; **Published:** April 5, 2013

**Copyright:** © 2013 Sun et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by Knowledge Innovation Project of Chinese Academy of Sciences (KSCX2-EW-J-20), National Natural Science Foundation of China grant (31070191) and U.S. National Science Foundation grant (ER-0431266). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: lijq@wbcas.cn (JQL); hcwang@wbcas.cn (HCW)

## Introduction

The eudicot order Trochodendrales [1] contains only two extant genera, both of which are monotypic: *Trochodendron* Sieb. & Zucc. and *Tetracentron* Oliver. Historically, these two genera have been treated either as the separate families Trochodendraceae and Tetracentraceae, or as the combined family Trochodendraceae [1–7]. The Trochodendraceae *sensu* APG III [1] appear to have been widespread in the Northern Hemisphere during the Paleogene and Neogene [7–15]. However, the two extant species of the family have small geographic ranges and are restricted to eastern Asia [16]. *Trochodendron aralioides* Sieb. & Zucc. is a large, evergreen shrub or small tree native to the mountains of Japan to South Korea and Taiwan, and the Ryukyu Islands [2,17], whereas *Tetracentron sinense* Oliver is a deciduous tree occurring in

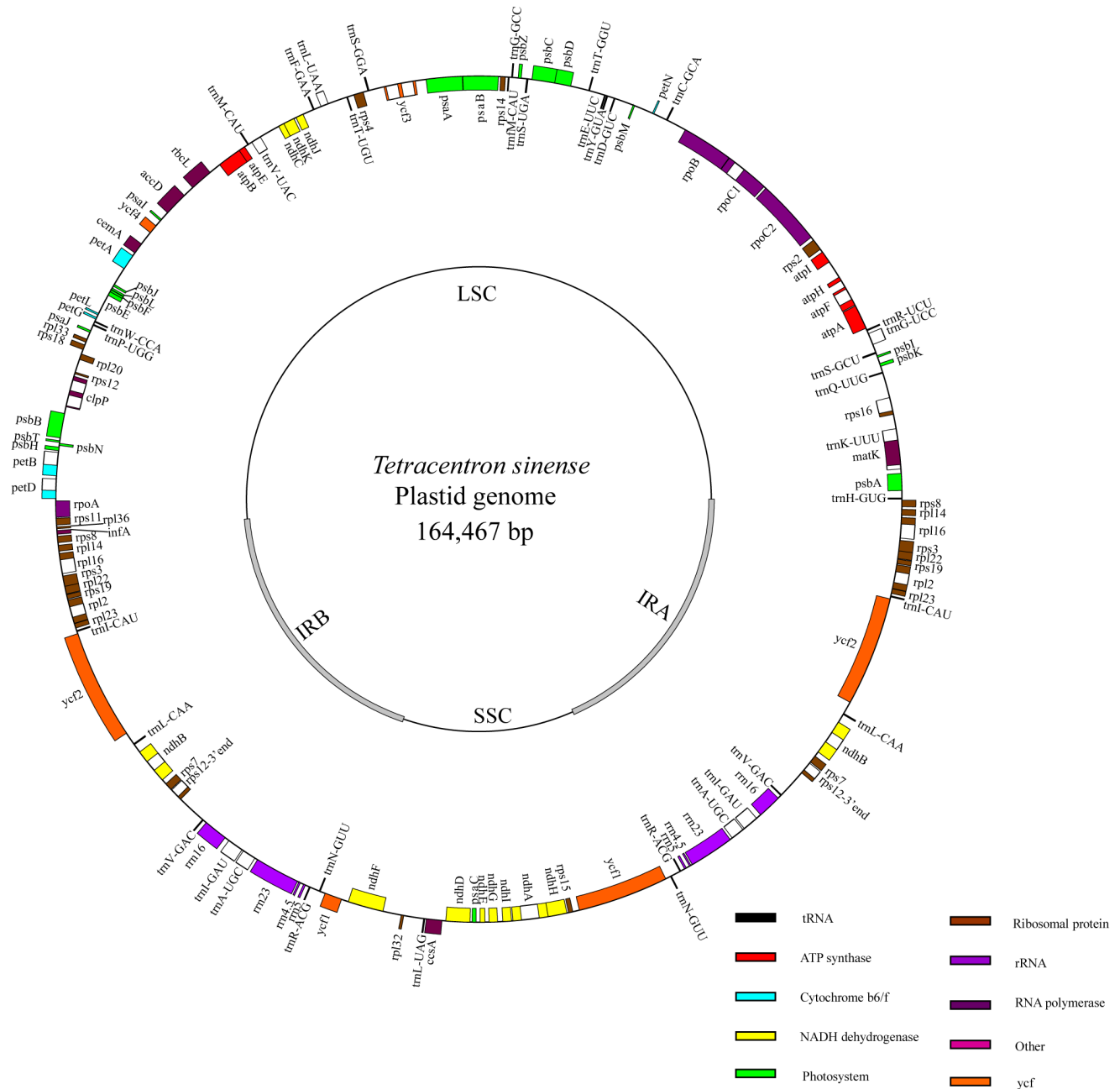
southwestern and central China and the eastern Himalayan regions. Both species are characterized by apetalous flowers arranged in cymose inflorescences and by loculicidal capsules that dehisce to release winged seeds [2,5,7,18]. Although earlier researchers reported that wood of Trochodendrales wood lacked vessels and thus suggested that Trochodendrales were among the earliest-diverging angiosperms, recent research has documented the presence of vessels in the wood of both genera [2,7,19].

Molecular phylogenetic studies, including analyses of complete plastid genome sequences, have routinely recovered Trochodendrales as an early-diverging member of the clade *Eudicotyledoneae* (*sensu* [20]; all italicized clade names follow this system), specifically as part of a strongly supported clade with Buxales and *Gunneridae*, or core eudicots [21–27]. However, the relationships among Trochodendrales, Buxales, and *Gunneridae* have often been only

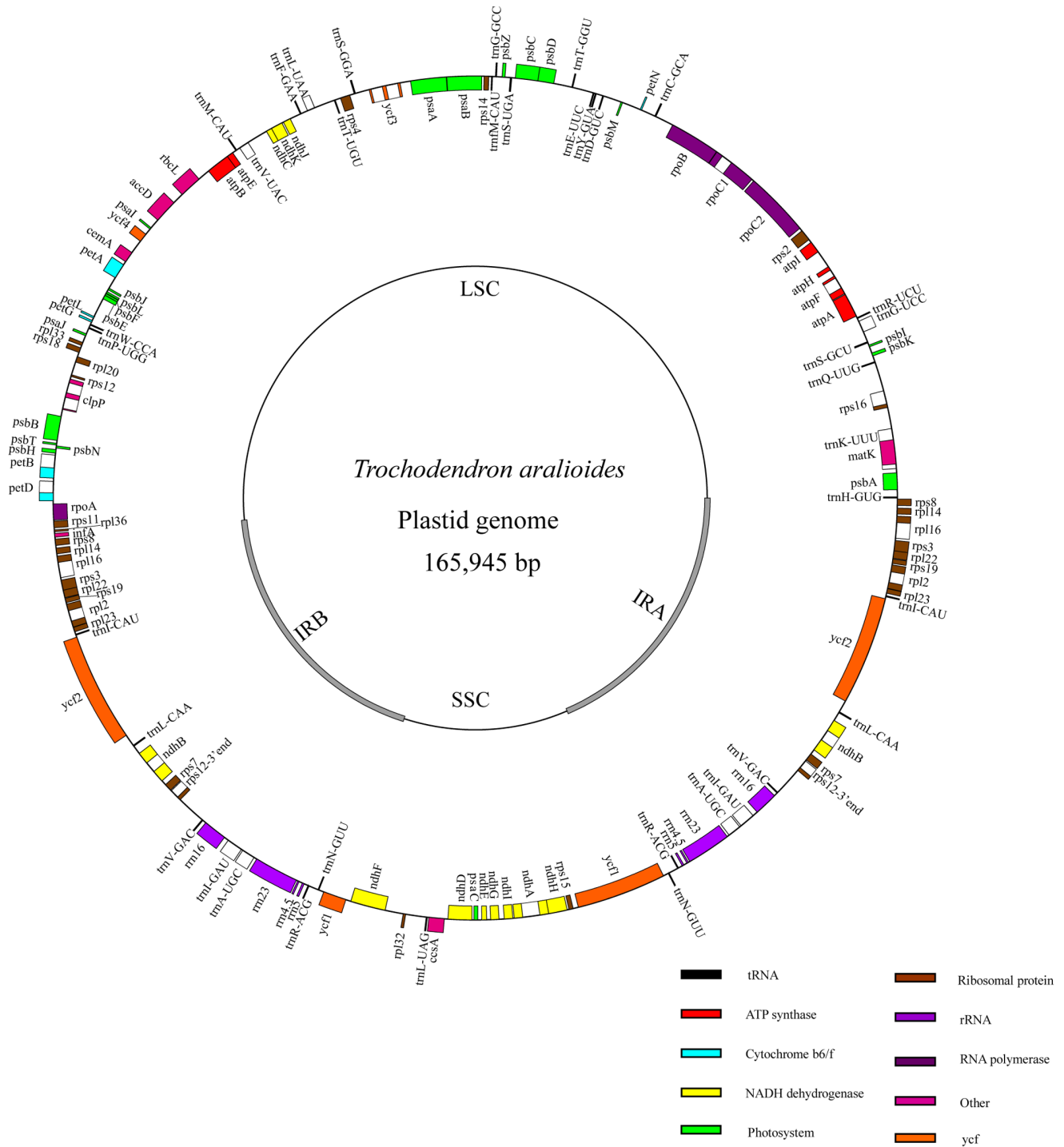
weakly supported. In the 17-gene analysis of Soltis et al. [28], which included data from all three plant genomes, Trochodendrales and Buxales were subsequent sisters to *Gunneridae*, with 100% and 98% BS support, respectively. However, other studies have found Buxales to be sister to *Gunneridae* with only weak support [24,26,29–30], whereas in other analyses Trochodendrales have appeared as sister to *Gunneridae* [27,31–32].

Complete plastid genome sequences have been used increasingly over the past decade to resolve deep-level phylogenetic relationships that have been unclear based on only a few genes. For example, recent plastid phylogenomic studies have helped to resolve key relationships among the earliest-diverging *Mesangiospermae* [33] as well as early-diverging *Eudicotyledoneae* and

*Pentapetalae* [26,34]. Indeed, the plastid genome represents an excellent source of characters for plant phylogenetics due to the generally strong conservation of plastid genome structure and its mix of sequence regions that vary tremendously in evolutionary rate [35–37], which enable plastid genome sequence data to be applied to phylogenetic problems at almost any taxonomic level in plants [26,38–43]. It is now relatively inexpensive to generate complete plastid genome sequence due to rapid improvements in next-generation sequencing (NGS) technologies [25,44–45] and due to the relatively small size of the plastid genome (~150 kb) and its structural conservation, which enable dozens of plastomes to be multiplexed per sequencing lane and facilitate relatively straightforward genome assembly [45–48].



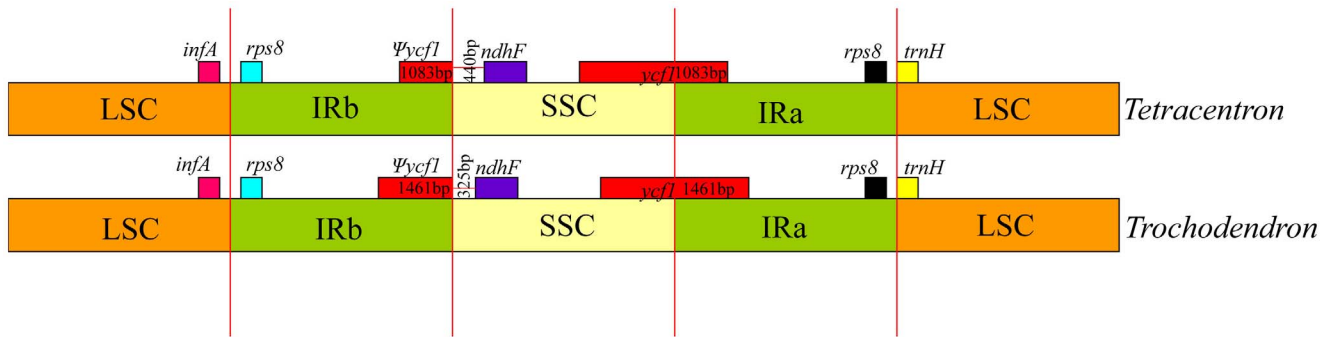
**Figure 1. Map of the *Tetracentron sinense* plastid genome.**  
doi:10.1371/journal.pone.0060429.g001



**Figure 2. Map of the *Trochodendron aralioides* plastid genome.**  
doi:10.1371/journal.pone.0060429.g002

Despite the promise of NGS technology for plastid genomics, the complete plastomes of only eight genera of early-diverging eudicots have been reported: *Ranunculus* (Ranunculaceae, Ranunculales), *Megaleranthis* (Ranunculaceae, Ranunculales), *Nandina* (Berberidaceae, Ranunculales), *Nelumbo* (Nelumbonaceae, Proteales), *Platanus* (Platanaceae, Proteales), *Meliosma* (Sabiaceae, Sabiales), *Trochodendron* (Trochodendraceae, Trochodendrales) and *Buxus* (Buxaceae, Buxales). Previous phylogenetic analyses

based on some of these complete genomes have not fully resolved the relationships among early-diverging eudicots, however; in addition to the uncertainty surrounding relationships of Buxales, Trochodendrales, and *Gunneridae*, the positions of Sabiales and Proteales remain poorly supported [26–27]. Plastome taxon sampling is still sparse in these clades, however, and additional sampling may help elucidate these recalcitrant relationships.



**Figure 3. Comparison of the IR junctions in *Tetracentron* and *Trochodendron*.**  
doi:10.1371/journal.pone.0060429.g003

In addition to their important role in phylogenetics, plastid genomes may be rich sources of population-level data. The non-recombination and uniparental inheritance of most plastid genomes can make plastid genomes extremely useful for population genetics, particularly for tracing maternal lineages [49–50]. For example, chloroplast simple sequence repeats (cpSSR) have been widely used in plant population genetics [51], including within early-diverging eudicots, where numerous cpSSR loci have been reported from the plastid genome of the endangered species *Megaleranthis saniculifolia* (Ranunculaceae) [52].

Here we report the complete plastid genome sequences of *Tetracentron sinense* and *Trochodendron aralioides* (the protein-coding and rRNA genes of *Trochodendron* cp genome were used for phylogenetic analyses in Moore et al. [26], but the cp genome structure of this genus has never been reported), as well as the results of new phylogenetic analyses based on adding *Tetracentron* and *Megaleranthis* genomes [52] to the 83-gene data set of Moore et al. [26]. We also compare the plastid genome structure of *Trochodendron* and *Tetracentron*, including the characterization of a significant expansion of the inverted repeat in both taxa, and we estimate the divergence time between the two genera. Finally, we characterize the distribution and location of cpSSRs in both *Tetracentron sinense* and *Trochodendron aralioides*, which provided further opportunity to study the population genetic structures of these two ancient relict species.

## Results

### Sequencing and Genome Assembly

Illumina paired-end sequencing produced 892.11 Mb of data for *Tetracentron sinense*. We obtained 9912310 raw reads of 90 bp in length. The N50 of contigs was 13,981 bp and the summed length of contigs was 143,709 bp. The mean coverage of this genome was 5424.2×. After de novo and reference-guided assembly, we obtained a cp genome containing nine gaps. PCR and Sanger sequencing were used for filling the gaps. Four junction regions between IRs and SSC/LSC were first determined based on de novo contigs, and subsequently confirmed by PCR amplifications and Sanger sequencing, sequenced results were compared with the assembled genome directly and no mismatch or indel was observed, which validated the accuracy of our assembly. The genome sequences of *Tetracentron sinense* and *Trochodendron aralioides* have been submitted to GenBank (GenBank IDs: KC608752 and KC608753).

### General Features of the *Tetracentron* and *Trochodendron* Plastomes

The plastid genome size of *Tetracentron sinense* is 164,467 base pairs (bp) (Figure 1), and that of *Trochodendron aralioides* is 165,945 bp (Figure 2). Both genomes show typical quadripartite structure, consisting of two copies of an inverted repeat (IR) separated by the large single-copy (LSC) and small single-copy

**Table 1.** Basic characteristic of the *Tetracentron sinense* and *Trochodendron aralioides* plastid genomes.

	<i>Tetracentron</i>	<i>Trochodendron</i>
total genome length	164467	165945
IR length	30231	30744
SSC length	19539	18974
LSC length	84466	85483
total length of coding sequence	94699	95168
total length of noncoding sequence	69768	70777
overall G/C content	38.1%	38.0%

All values given are in base pairs (bp), unless otherwise noted.  
doi:10.1371/journal.pone.0060429.t001

**Table 2.** The principal noncoding regions contributing to the size difference between the *Tetracentron* and *Trochodendron* plastid genomes.

Spacer region or intron names	<i>Tetracentron</i>	<i>Trochodendron</i>	length difference
<i>trnK-UUU/rps16</i> spacer	870	1308	438
<i>rps16/trnQ-UUG</i> spacer	1529	1797	268
<i>trnS-GCU/trnG-UCC</i> spacer	505	658	153
<i>trnE-UUC/trnT-GGU</i> spacer	957	1316	359
<i>trnT-UGU/trnL-UAA</i> spacer	1199	1309	110
<i>petA/psbJ</i> spacer	1146	754	−392
<i>ycf1/ndhF</i> spacer	440	325	−115
<i>*rpl16</i> intron	865	972	107

All sizes are in base pairs. The only locus residing in the IR is marked with an asterisk (\*).

doi:10.1371/journal.pone.0060429.t002





**Table 3.** List of genes present in the plastid genomes of *Tetracentron sinense* and *Trochodendron aralioides*.

	Group of genes	Name of genes
Protein synthesis and DNA replication	Ribosomal RNAs	<i>rrn4.5</i> (×2) <i>rrn5</i> (×2) <i>rrn16</i> (×2) <i>rrn23</i> (×2)
	Transfer RNAs	<i>trnH-GUG</i> <i>trnK-UUU</i> * <i>trnQ-UUG</i> <i>trnS-GCU</i> <i>trnG-UCC</i> * <i>trnR-UCU</i> <i>trnC-GCA</i> <i>trnD-GUC</i> <i>trnY-GUA</i> <i>trnE-UUC</i> <i>trnT-GGU</i> <i>trnS-UGA</i> <i>trnG-GCC</i> <i>trnM-CAU</i> <i>trnS-GGA</i> <i>trnT-UGU</i> <i>trnL-UAA</i> * <i>trnF-GAA</i> <i>trnV-UAC</i> * <i>trnM-CAU</i> <i>trnW-CCA</i> <i>trnP-UGG</i> <i>trnI-GAU</i> * (×2) <i>trnL-CAA</i> (×2) <i>trnV-GAC</i> (×2) <i>trnI-GAU</i> (×2) <i>trnA-UGC</i> * (×2) <i>trnR-ACG</i> (×2) <i>trnN-GUU</i> (×2) <i>trnL-UAG</i>
	small subunit	<i>rps2</i> <i>rps3</i> <i>rps4</i> <i>rps7</i> (×2) <i>rps8</i> <i>rps11</i> <i>rps12</i> * (×2) <i>rps14</i> <i>rps15</i> <i>rps16</i> * <i>rps18</i> <i>rps19</i>
	Ribosomal proteins large subunit	<i>rpl2</i> * (×2) <i>rpl14</i> <i>rpl16</i> * <i>rpl20</i> <i>rpl22</i> <i>rpl23</i> (×2) <i>rpl32</i> <i>rpl33</i> <i>rpl36</i>
	RNA polymerase	<i>rpoA</i> <i>rpoB</i> <i>rpoC1</i> * <i>rpoC2</i>
	Photosynthesis	Photosystem I
Photosystem II		<i>psbA</i> <i>psbB</i> <i>psbC</i> <i>psbD</i> <i>psbE</i> <i>psbF</i> <i>psbH</i> <i>psbI</i> <i>psbJ</i> <i>psbK</i> <i>psbL</i> <i>psbM</i> <i>psbN</i> <i>psbT</i> <i>psbZ</i>
Cytochrome b6/f		<i>petA</i> <i>petB</i> * <i>petD</i> * <i>petG</i> <i>petL</i> <i>petN</i>
ATP synthase		<i>atpA</i> <i>atpB</i> <i>atpE</i> <i>atpF</i> * <i>atpH</i> <i>atpI</i>
NADH dehydrogenase		<i>ndhA</i> * <i>ndhB</i> * (×2) <i>ndhC</i> <i>ndhD</i> <i>ndhE</i> <i>ndhF</i> <i>ndhG</i> <i>ndhH</i> <i>ndhI</i> <i>ndhJ</i> <i>ndhK</i>
Large subunit of Rubisco		<i>rbcL</i>
Miscellaneous proteins		Subunit of Acetyl-CoA-carboxylase
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Envelope membrane protein	<i>cemA</i>
	Protease	<i>clpP</i> *
	Translational initiation factor	<i>infA</i>
	Maturase	<i>matK</i>
Genes of unknown function	Hypothetical conserved coding frame	<i>ycf1</i> <i>ycf2</i> (×2) <i>ycf3</i> * <i>ycf4</i>

Genes with introns are marked with asterisks (\*).

doi:10.1371/journal.pone.0060429.t003

(SSC) regions (Table 1). The IR exhibits a significant expansion relative to most other angiosperms at the LSC/IR junction; specifically, the IR in both *Tetracentron* and *Trochodendron* has expanded to include the entirety of the *rps19*, *rpl22*, *rps3*, *rpl16*, *rpl14*, and *rps3* genes (Figures 1, 2). The SSC/IR boundary occurs within the *ycf1* gene, as is typical in angiosperms, but is slightly expanded in the *Trochodendron* genome to include 1461 bp of the 5' end of *ycf1* (versus 1083 bp in *Tetracentron*; Figure 3). This expansion of the IR at the SSC junction contributes to the difference in length between the two Trochodendrales plastomes; the remainder of the difference is largely the result of length differences among various noncoding regions (Table 2).

Both genomes contain 119 genes (79 protein-coding genes, 30 tRNA genes, and 4 rRNA genes) arranged in the same order, of which 24 are duplicated in the IR regions (Table 3). Sequence divergence between *Tetracentron* and *Trochodendron* in coding regions is low (Table 4, Figures 4, 5). Only 7 genes (*rps11*, *rpoA*, *rpl32*, *rps16*, *ndhF*, *ycf1*, and *rpl36*) exhibit divergences of more than 2%, and 12 genes have an identical sequence (Table 4, Figure 4). The genes *ndhF*, *ycf1*, and *rpl36* have the highest sequence divergences (2.7%, 3.5% and 4.4%, respectively). The coding regions account for 57.5% and 57.3% of the *Tetracentron* and *Trochodendron* plastid genomes, respectively. For both cp genomes, single introns are present in 18 genes, whereas three genes (*rps12*, *clpP*, and *ycf3*) have two introns (Table 5). The overall genomic G/C nucleotide composition is 38.1% and 38.0% for *Tetracentron* and *Trochodendron*, respectively; detailed A/T contents of different regions of the plastome for both genomes are listed in Table 6. Due to the lower A/T content of the four rRNA genes, the IR regions possess lower A/T content than the single-copy regions.

### Characterization of SSR Loci

In all, 154 SSR loci (77 each from *Tetracentron sinense* and *Trochodendron aralioides*) were detected in the two plastid genomes, of which 123 are mononucleotide repeats, 28 are dinucleotide repeats, two are trinucleotide repeats, and one is a tetranucleotide repeat (Table 7). Nearly all of the SSR loci are composed of A/T repeats (Table 7), and these SSR loci are mostly present in noncoding regions. The tetranucleotide locus identified in *Tetracentron* is in the first intron of *ycf3*. The two trinucleotide loci in *Trochodendron* are both located in the spacer region between *trnK-UUU* and *rps16*. The unique C mononucleotide repeat from *Trochodendron* is present in the *trnV-ndhC* intergenic spacer region.

### Phylogenetic and Molecular Dating Analyses

ML analyses of the 83-gene, 88-taxon data set yielded a tree with a similar topology and bootstrap support (BS) values (Figure 6) as that of the plastid phylogenomic study of Moore et al. [26]. The clades of *Trochodendron*+*Tetracentron* and *Ranunculus*+*Megaleranthis* were supported with 100% ML BS support. Trochodendrales are sister to the remaining angiosperms with high support (BS = 100%), but Buxaceae are sister to Gunneridae with only 67% BS support.

Molecular dating analyses suggest that *Trochodendron* and *Tetracentron* diverged between 44–30 million ago. The crown group 95% highest posterior density (HPD) age estimates for other major lineages of *Pentapetalae* were as follows: *Superasteridae* (115–109 mya), *Dilleniaceae*+*Superrosidae* (116–112 mya), *Superrosidae* (114–111 mya), *Santalales* (98–75 mya), *Caryophyllales* (76–60 mya), *Asteridae* (104–99 mya), *Rosidae* (111–108 mya), *Vitaceae*+*Saxifragales* (114–110 mya), and *Saxifragales* (109–107 mya).



**Table 4.** Comparisons of the protein-coding genes of *Tetracentron* and *Trochodendron*.

Gene	Length in <i>Tetracentron</i>	Length in <i>Trochodendron</i>	Number of nucleotide differences	Proportion of nucleotide differences	Number of indel differences
<i>petL</i>	102	102	0	0	0
<i>psaI</i>	111	111	0	0	0
<i>psaJ</i>	129	129	0	0	0
<i>psbE</i>	252	252	0	0	0
<i>psbF</i>	120	120	0	0	0
<i>psbJ</i>	123	123	0	0	0
<i>psbL</i>	117	117	0	0	0
<i>psbT</i>	108	108	0	0	0
<i>rpl23</i>	288	288	0	0	0
<i>rps19</i>	279	279	0	0	0
<i>rps7</i>	468	468	0	0	0
<i>rps8</i>	399	399	0	0	0
<i>rpl2</i>	825	825	1	0.00121	0
<i>rps3</i>	657	657	1	0.00152	0
<i>petD</i>	504	504	1	0.00198	0
<i>rpl16</i>	501	501	1	0.00249	0
<i>rpl14</i>	369	369	1	0.00271	0
<i>ycf2</i>	6879	6897	19	0.00276	1
<i>ndhB</i>	1533	1533	5	0.00326	0
<i>ycf3</i>	507	507	2	0.00394	0
<i>rpl33</i>	201	201	1	0.00498	0
<i>psbZ</i>	189	189	1	0.00529	0
<i>psaA</i>	2253	2253	12	0.00533	0
<i>psbK</i>	186	186	1	0.00538	0
<i>rps12</i>	372	372	2	0.00538	0
<i>psbA</i>	1062	1062	6	0.00565	0
<i>rpl20</i>	354	354	2	0.00565	0
<i>rpoC1</i>	2049	2070	12	0.00586	1
<i>atpA</i>	1524	1524	9	0.00591	0
<i>rpl22</i>	486	480	3	0.00625	1
<i>ndhJ</i>	477	477	3	0.00629	0
<i>psbD</i>	1062	1062	7	0.00659	0
<i>petA</i>	963	963	7	0.00727	0
<i>rpoB</i>	3213	3213	24	0.00747	0
<i>psbN</i>	132	132	1	0.00758	0
<i>psaB</i>	2205	2205	17	0.00771	0
<i>psbC</i>	1422	1422	11	0.00774	0
<i>atpH</i>	246	246	2	0.00813	0
<i>psaC</i>	246	246	2	0.00813	0
<i>ndhA</i>	1095	1095	9	0.00822	0
<i>rps4</i>	606	606	5	0.00825	0
<i>infA</i>	234	234	2	0.00855	0
<i>atpB</i>	1497	1497	13	0.00868	0
<i>cemA</i>	690	690	6	0.0087	0
<i>petG</i>	114	114	1	0.00877	0
<i>psbI</i>	111	111	1	0.00901	0
<i>rbcL</i>	1428	1428	13	0.0091	0
<i>petB</i>	648	648	6	0.00926	0
<i>atpI</i>	744	744	7	0.00941	0

**Table 4.** Cont.

Gene	Length in <i>Tetracentron</i>	Length in <i>Trochodendron</i>	Number of nucleotide differences	Proportion of nucleotide differences	Number of indel differences
<i>clpP</i>	609	609	6	0.00985	0
<i>rps14</i>	303	303	3	0.0099	0
<i>atpE</i>	402	402	4	0.00995	0
<i>ccsA</i>	966	966	10	0.01035	0
<i>psbB</i>	1527	1527	16	0.01048	0
<i>accD</i>	1491	1491	16	0.01073	0
<i>ndhK</i>	822	858	9	0.01095	1
<i>ndhC</i>	363	363	4	0.01102	0
<i>petN</i>	90	90	1	0.01111	0
<i>ndhG</i>	531	531	6	0.0113	0
<i>rpoC2</i>	4137	4146	50	0.01209	1
<i>ndhD</i>	1503	1503	18	0.01264	0
<i>rps2</i>	711	711	9	0.01266	0
<i>psbH</i>	222	222	3	0.01351	0
<i>ndhI</i>	543	543	8	0.01473	0
<i>atpF</i>	555	555	9	0.01622	0
<i>matK</i>	1536	1536	25	0.01628	0
<i>ndhE</i>	306	303	5	0.0165	1
<i>rps18</i>	303	303	5	0.0165	0
<i>ndhH</i>	1182	1182	20	0.01692	0
<i>ycf4</i>	555	555	10	0.01805	0
<i>rps15</i>	273	273	5	0.01832	0
<i>psbM</i>	105	105	2	0.01905	0
<i>rps11</i>	417	417	9	0.02158	0
<i>rpoA</i>	1014	1014	24	0.02367	0
<i>rpl32</i>	162	162	4	0.02469	0
<i>rps16</i>	227	227	6	0.02622	0
<i>ndhF</i>	2223	2223	61	0.02744	0
<i>ycf1</i>	5688	5691	195	0.0345	6
<i>rpl36</i>	114	114	5	0.04386	0

Genes are ranked from lowest to highest proportion of nucleotide differences.  
doi:10.1371/journal.pone.0060429.t004

## Discussion

### Expansion of the IR Region in Trochodendrales Plastomes

The plastid genomes of *Tetracentron* and *Trochodendron* exhibit the typical gene content and genome structure of angiosperms [37,53–54], with the notable exception of a significantly expanded IR region (Figures 1, 2, 3). This ~4 kb expansion is responsible for the relatively large size of both Trochodendrales plastomes, which are ~4–5 kb larger than the typical upper size range of angiosperm plastid genomes, including those of nearly all other early-diverging eudicots (Table 8). Significant expansion, contraction, and even loss of the IR appears to be an evolutionarily uncommon phenomena but are nonetheless associated with much of the more significant variation in plastome size in angiosperms. For example, the largest known angiosperm plastome, that of *Pelargonium x hortorum*, also possesses the largest known IR, at ~76 kb in length [55]. Other significant IR expansions and contractions have been found in Campanulaceae [56–57], Apiaceae [58], and *Lemna* (Araceae) [59].

### Impact of Additional Taxon Sampling on Basal Eudicot Phylogeny

The inclusion of *Megaleranthis* and *Tetracentron* in our analyses had no effect on the relationships among the major early-diverging eudicot lineages, and very little effect on support values. Of the basal splits among the eudicots with BS values less than 100% in both the current tree and that of Moore et al. [26], all were within 3% BS value. For example, the sister relationship of Buxales and *Gunneridae* is 70% in Moore et al. [26] vs. 67% with the inclusion of *Megaleranthis* and *Tetracentron*, and the sister relationship of Sabiales and Proteales has BS support of 80% in Moore et al. [26] vs. 83% in the current analyses. These similar values are unsurprising given that *Tetracentron* and *Trochodendron* are found to be relatively closely related in our analyses. Indeed, the relatively low sequence divergence between the *Tetracentron* and *Trochodendron* plastid genomes supports the taxonomic placement of Tetracentraceae within Trochodendraceae, as advocated by APG III [1]. Although it is possible that the addition of the noncoding regions of the

**Table 5.** Exon and intron lengths (bp) in plastid genes containing introns in *Tetracentron sinense* and *Trochodendron aralioides*, respectively.

Gene	Exon 1 (Te/Tr)	Intron 1 (Te/Tr)	Exon 2 (Te/Tr)	Intron 2 (Te/Tr)	Exon 3 (Te/Tr)
<i>trnK-UUU</i>	37/37		35/35		
<i>trnG-UCC</i>	24/24	698/698	48/48		
<i>trnL-UAA</i>	35/35	444/442	50/50		
<i>trnV-UAC</i>	39/39	583/585	37/37		
<i>trnI-GAU</i>	42/42	954/954	35/35		
<i>trnA-UGC</i>	38/38	794/794	35/35		
<i>petB</i>	6/6	793/797	642/642		
<i>petD</i>	8/8	704/709	496/496		
<i>atpF</i>	145/145	727/724	410/410		
<i>ndhA</i>	553/553	1106/1084	542/542		
<i>ndhB</i>	777/777	700/700	756/756		
<i>rpl2</i>	391/391	671/674	434/434		
<i>rpl16</i>	9/9	865/972	402/402		
<i>rps12</i>	114/114		232/232	538/536	26/26
<i>rpoC1</i>	432/432	728/714	1617/1638		
<i>clpP</i>	71/71	682/710	292/292	659/650	246/246
<i>ycf3</i>	124/124	734/725	230/230	731/758	153/153
<i>rps16</i>	40/40	831/844	227/227		

The *rps12* gene is trans-spliced, and hence the length of intron 1 is unknown.  
doi:10.1371/journal.pone.0060429.t005

plastid genome (or at least those noncoding regions that can be aligned) to our data set may improve support for these relationships, we may have to look to the other plant genomes for a confident resolution of relationships among the early-diverging eudicots. In fact, the sister relationship of Buxales and *Gunneridae* received high support (BS = 98%) in the 17-gene analyses of Soltis et al. [28], which employed a combination of 11 plastid genes, 18S and 26S nuclear rDNA, and 4 mitochondrial genes. However, the sister relationship of Sabiales and Proteales were more poorly supported (BS = 59%) in Soltis et al. [28].

#### Divergence Time Between *Tetracentron* and *Trochodendron*

Cenozoic Trochodendrales fossils are known throughout the Northern Hemisphere, with the Paleocene *Nordenskioldia* the earliest certain fossil of the order [7–15]. Both *Tetracentron* and *Trochodendron* had wide distributions in the Northern Hemisphere during the Paleogene and Neogene. Fossil remains of *Tetracentron* have been found in Japan [60–61], Idaho [62], Princeton, British

Columbia and Republic, Washington [63], and Iceland [15]; *Trochodendron* fossil remains have been reported from Kamchatka [64], Japan [11], Idaho and Oregon [11–12], Washington [7], and British Columbia [63]. Our estimate of the divergence time between the two genera of Trochodendraceae (44–30 mya) encompasses the recent estimate of 37–31 mya from Bell et al. [65], which was based on analysis of 567 taxa and three genes, as well as the mid-Eocene estimate of ~45 mya derived from the *rbcL* analysis of Anderson et al. [66], which employed numerous fossil constraints from the early-diverging eudicots. The congruence among these studies and with the fossil record suggests that a mid-to late Eocene divergence for the two extant Trochodendraceae lineages may be a reasonable estimate.

#### Analysis of Plastid SSR Loci in the Trochodendrales

Because microsatellite loci, including cpSSRs, often exhibit high variation within species, they are considered valuable molecular markers for population genetics [67–69]. A limited number of SSR loci were recently characterized for *Tetracentron* [70], but no cpSSR loci are available for Trochodendraceae. The 77 cpSSR loci that were identified in both *Tetracentron* and *Trochodendron* represent ~42% more loci than the 54 loci reported in the plastid genome of *Megaleranthis* (Ranunculaceae), the only other early-diverging eudicot for which a comprehensive analysis of cpSSR loci is available. The abundant and varied cpSSR loci identified in Trochodendrales will be useful in characterizing the population genetics of both extant species, which are of conservation interest in the wild because of their relatively narrow, presumably relictual distributions, and decreasing numbers [71]. *Tetracentron* is officially afforded second-class protection in China.

**Table 6.** A/T content (%) of different regions in *Tetracentron* and *Trochodendron*.

Region	<i>Tetracentron</i>	<i>Trochodendron</i>
overall	61.86	61.98
LSC	63.50	63.74
IR	57.63	57.83
SSC	67.84	67.48
Protein-coding regions	61.58	61.53

doi:10.1371/journal.pone.0060429.t006

**Table 7.** Distribution of SSR loci in the plastid genomes of *Tetracentron* and *Trochodendron*.

Base	Length	Position in plastid genome
SSR loci in <i>Tetracentron</i>		
A	10	2085–2094 7164–7173 9478–9487 17266–17275 39220–39229 47812–47821 58880–58889 69930–69939 124816–124825 136417–136426 141648–141657
	11	9611–9621 46892–46902 47147–47157 50813–50823 75797–75807 80873–80883 82302–82312 133069–133079 160432–160442
	12	217–228 49977–49988 50332–50343 118899–118910 162450–162461 163452–163463 163940–163951
	14	65157–65170
	15	38842–38856
	17	39891–39907
	18	74838–74855
	22	72886–72907
T	10	5266–5275 6724–6733 9153–9162 19332–19341 54468–54477 63461–63470 67706–67715 107277–107286 112508–112517 117373–117382 118300–118309 121204–121213 126456–126465 130614–130623
	11	7004–7014 7679–7689 13144–13154 31361–31371 37925–37935 47779–47789 67810–67820 76013–76023 88492–88502
	12	55307–55318 71723–71734 84983–84994 85471–85482 86473–86484 118884–118895 119027–119038
	13	13902–13914
	14	72926–72939
AT	10	1734–1743 20833–20842 50404–50413–63181–63190
	12	4862–4873 12996–13007 114822–114833
	14	60686–60699
TA	10	34083–34092 34111–34120 114741–114750
	14	49132–49145
TAAA	20	46875–46894
SSR loci in <i>Trochodendron</i>		
A	10	118854–118863 126258–126267 142993–143002 163821–163830 18142–18151 40389–40398 41060–41069 51091–51100 6136–6145 68969–68978 76681–76690 86529–86538
	11	134406–134416 16427–16437 30306–30316 39963–39973 51490–51500 70911–70921 81823–81833 9789–9799
	12	10420–10431 48058–48069 48322–48333
	13	164932–164944
	16	161805–161820 73777–73792 75726–75741
	15	46189–46203
	17	214–230 83299–83315 9304–9320
T	10	108427–108436 120424–120433 121028–121037 122665–122674 131951–131960 164891–164900 20189–20198 40375–40387 48933–48942 53154–53163 53339–53348 5700–5709 6030–6039 68604–68613 72934–72943 83282–83291 87599–87608
	11	127885–127895 14709–14719 55604–55614 57547–57557
	12	50271–50282
	13	73814–73826 86485–86497
	14	76896–76909
	15	48889–48903
	16	89609–89624
AT	10	1724–1733 51556–51565 64459–64468
	12	4921–4932 4943–4954 4984–4995 4998–5009 5044–5055 5085–5096 5099–5110 5145–5156 5186–5197 5200–5211
	18	73275–73292
TA	10	1738–1747 21689–21698
TAA	18	5016–5033 5218–5235
C	10	55999–56008

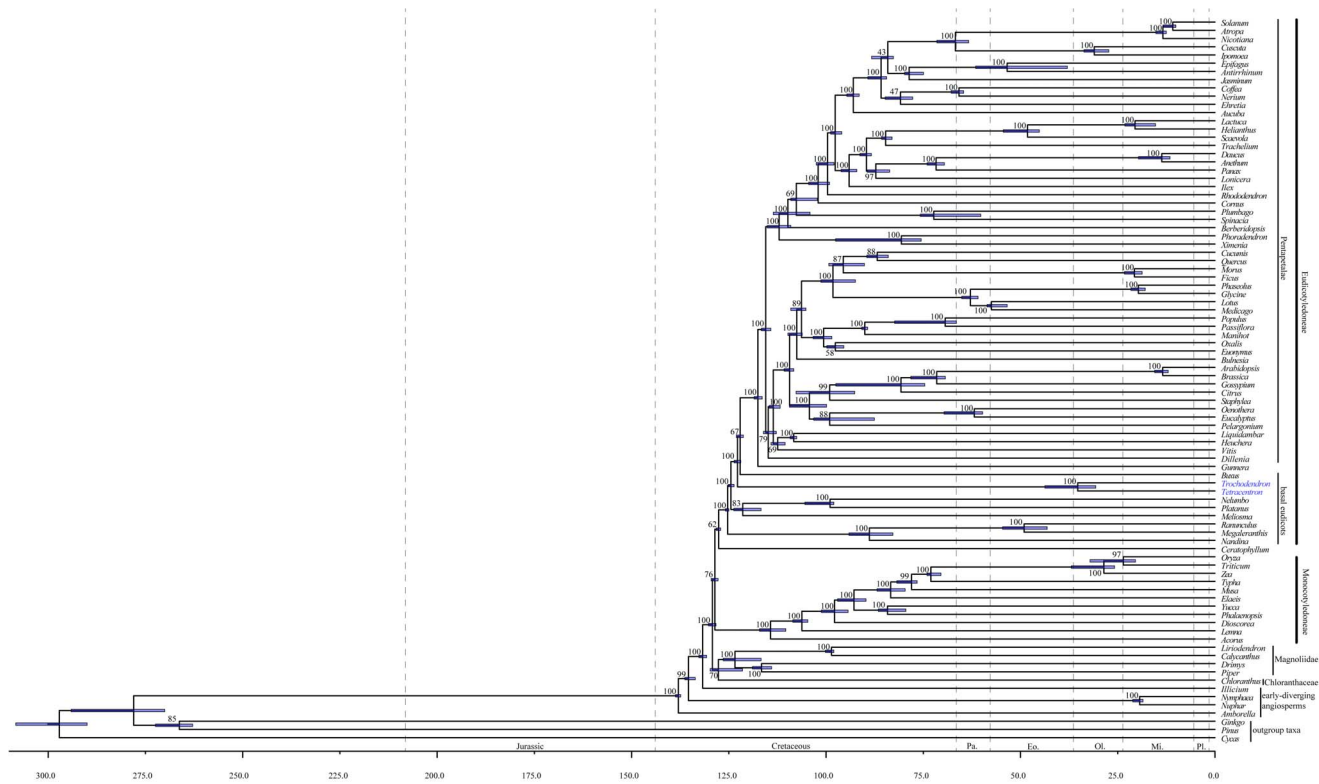
doi:10.1371/journal.pone.0060429.t007

## Materials and Methods

### Sample Preparation, Sequencing, and Assembly

Fresh leaves of *Tetracentron sinense* were collected from the Kunming Institute of Botany at the Chinese Academy of Sciences,

and a voucher was deposited at the Herbarium of Wuhan Botanical Garden, Chinese Academy of Science (HIB). Chloroplast DNA was isolated following the protocol of Zhang et al. [45], and an Illumina library was constructed following the manufacturer's protocol (Illumina). The DNA was indexed by tag and



**Figure 6. A maximum likelihood tree determined by GARLI ( $-\ln L = -1095466.026$ ) for the 83-gene, 88-taxon data set.** Numbers associated with branches are ML bootstrap support values. Error bars around nodes correspond to 95% highest posterior distributions of divergence times based on 6 fossils using the program BEAST. Eo = Eocene, Mi = Miocene, Ol. = Oligocene, Pa = Paleocene, Pl = Pliocene. doi:10.1371/journal.pone.0060429.g006

sequenced together with eight other species in one lane of an Illumina Genome Analyzer Iix at Beijing Genomics Institute (BGI) in Shenzhen, China. Illumina Pipeline 1.3.2 was used conducting image analysis and base calling. Raw sequence reads produced by Illumina paired-end sequencing were filtered for high quality reads which were subsequently assembled into contigs with a minimum length of 100 bp using SOAPdenovo [72] with the Kmer = 57. Contigs were aligned to the *Trochodendron aralioides* plastid genome using BLAST (<http://blast.ncbi.nlm.nih.gov/>), and aligned contigs were ordered according to the reference genome.

**Genome Annotation and Analysis**

The *Tetracentron* and *Trochodendron* plastid genomes were annotated with DOGMA [73] and BLAST tools from NCBI (the National Center for Biotechnology Information). Physical maps were generated using GenomeVx [74] with subsequent manual editing. Sequence divergence between the *Tetracentron* and *Trochodendron* plastid genomes was evaluated using DnaSP version 5.10 [75], and genome sequence identity plots were generated using mVISTA [76] (<http://genome.lbl.gov/vista/mvista/submit.shtml>). Msatfinder ver. 1.6.8 [77] was used to identify SSR loci by manually setting repeat units.

**Table 8. Numbers of genes (including genes that span IR/SC junctions) in the IR regions of early-diverging eudicots.**

Basal eudicot lineages	Species	Genes in IR region	cp genome size (bp)
Ranunculales	<i>Ranunculus macranthus</i>	20	155129
	<i>Megaleranthis saniculifolia</i>	19	159924
	<i>Nandina domestica</i>	19	156599
Proteales	<i>Nelumbo lutea</i>	18	163206
	<i>Platanus occidentalis</i>	19	161791
Sabiales	<i>Meliosma</i> aff. <i>cuneifolia</i>	18	160357
Buxales	<i>Buxus microphylla</i>	18	159010
Trochodendrales	<i>Tetracentron sinense</i>	24	164467
	<i>Trochodendron aralioides</i>	24	165945

doi:10.1371/journal.pone.0060429.t008

## Phylogenetic and Divergence Time Analyses

All protein-coding sequences, as well as all rRNA sequences, were extracted from the *Tetracentron* and *Megaleranthis* plastome [52] and added manually to the 83-gene, 86-taxon alignment of Moore et al. [26]. ML analyses were performed on the concatenated 83-gene data set using the following partitioning strategy: (1) codon positions 1 and 2 together; (2) codon position 3; and (3) rRNA genes. The optimal nucleotide sequence model was selected for each partition using jModelTest 2.1.1 using the Decision Theory (DT) criterion [78]. The following models were selected: TVM+I+ $\Gamma$  for codon positions 1+2 and for codon position 3, and TIM1+ I+ $\Gamma$  for rRNA.

Partitioned ML analyses were conducted using GARLI 2.0 [79]. A total of ten search replicates were conducted to find the optimal tree, and nonparametric bootstrap support was assessed with 100 replicates [80]. All ML searches used random taxon addition to build starting trees.

Divergence times were estimated using BEAST version 1.7.4 [81], using the same dating strategies employed in Moore et al.

## References

- Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn Soc* 161: 105–121.
- Smith AC (1945) A taxonomic review of *Trochodendron* and *Tetracentron*. *J Arnold Arboretum Harvard University* 26: 123–142.
- Cronquist A (1981) *An Integrated System of Classification of Flowering Plants*. New York: Columbia University Press.
- Endress PK (1986) Reproductive structures and phylogenetic significance of extant primitive angiosperms. *Plant Syst Evol* 152: 1–28.
- Endress PK, Igersheim A (1999) Gynoecium diversity and systematics of the basal eudicots. *Bot J Linn Soc* 130: 305–393.
- Magallón S, Crane PR, Herendeen PS (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Ann Missouri Bot Gard* 86: 297–372.
- Pigg KB, Wehr WC, Ickert-Bond SM (2001) *Trochodendron* and *Nordenskiöldia* (Trochodendraceae) from the Middle Eocene of Washington State, U.S.A. *Int J Plant Sci* 162: 1187–1198.
- Crane PR (1989) Paleobotanical evidence on the early radiation of nonmagnoliid dicotyledons. *Plant Syst Evol* 162: 165–191.
- Crane PR, Manchester SR, Dilcher DL (1990) A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota. *Fieldiana Geol NS* 20: 1–63.
- Crane PR, Manchester SR, Dilcher DL (1991) Reproductive and vegetative structure of *Nordenskiöldia* (Trochodendraceae), a vesselless dicotyledon from the early Tertiary of the Northern Hemisphere. *Am J Bot* 8: 1311–1334.
- Manchester SR, Crane PR, Dilcher DL (1991) *Nordenskiöldia* and *Trochodendron* fruits (Trochodendraceae) from the Miocene of northwestern North America. *Bot Gaz* 152: 357–368.
- Fields PF (1996a) The Succor Creek flora of the middle Miocene Sucker Creek Formation, southwestern Idaho and eastern Oregon: systematics and paleoecology. PhD diss. Michigan State University, East Lansing.
- Fields PF (1996b) A *Trochodendron* infructescence from the 15 Ma Succor Creek flora in Oregon: a geographic and possibly temporal range extension. *Am J Bot* 83(suppl): 110.
- Manchester SR (1999) Biogeographical relationships of North American Tertiary floras. *Ann Missouri Bot Gard* 86: 472–522.
- Grimsson F, Denk T, Zetter R (2008) Pollen, fruits, and leaves of *Tetracentron* (Trochodendraceae) from the Cainozoic of Iceland and western North America and their palaeobiogeographic implications. *Grana* 47: 1–14.
- Watson L, Dallwitz MJ (2006) *The families of flowering plants: descriptions, illustrations, identification, information retrieval*. Version 3.
- Mabberley DJ (1987) *The plant-book*. Cambridge: Cambridge University Press.
- Doweld AB (1998) Carpology, seed anatomy and taxonomic relationships of *Tetracentron* (Tetracentraceae) and *Trochodendron* (Trochodendraceae). *Ann Bot* 82: 413–443.
- Li HF, Chaw SM, Du CM, Ren Y (2011) Vessel elements present in the secondary xylem of *Trochodendron* and *Tetracentron* (Trochodendraceae). *Flora* 206: 595–600.
- Cantino PD, Doyle JA, Graham SW, Judd WS, Olmstead RG et al. (2007) Towards a phylogenetic nomenclature of *Tracheophyta*. *Taxon* 56: 822–846.
- Soltis DE, Soltis PS, Endress PK, Chase MW (2005) Phylogeny and Evolution of the Angiosperms. Sunderland, MA: Sinauer.
- Qiu Y, Dombrowska O, Lee J, Li L, Whitlock BA, et al. (2005) Phylogenetic analysis of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *Int J Plant Sci* 166: 815–842.

[26]. In addition to the three calibration points (used in Moore et al. [26]) of minimum ages of 131.8 mya for angiosperms [82–85], 125 mya for eudicots [83,86], and 85 mya for the most recent common ancestor of *Quercus* and *Cucumis* [26], we additionally constrained the stem lineage of Malpighiales using a minimum of 89.3 my [87] and the node uniting *Cabycanthus* and *Liriodendron* using 98 my [88], and set the age of Proteales to a minimum of 98 my [89].

## Acknowledgments

We thank the anonymous reviewers for their helpful comments on earlier versions of this manuscript.

## Author Contributions

Conceived and designed the experiments: JQL HCW. Performed the experiments: YXS MJM APM. Analyzed the data: YXS MJM. Contributed reagents/materials/analysis tools: YXS MJM JQL HCW. Wrote the paper: YXS MJM PSS DES HCW.

- Qiu YL, Li L, Hendry TA, Li R, Taylor DW, et al. (2006) Reconstructing the basal angiosperm phylogeny: evaluating information content of mitochondrial genes. *Taxon* 55: 837–856.
- Worberg A, Quandt D, Barniske A-M, Löhne C, Hilu KW, et al. (2007) Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Org Divers Evol* 7: 55–77.
- Soltis DE, Moore MJ, Burleigh JG, Bell CD, Soltis PS (2010) Assembling the Angiosperm Tree of Life: progress and future prospects. *Ann Missouri Bot Gard* 97: 514–526.
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc Natl Acad Sci USA* 107: 4623–4628.
- Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley M, et al. (2011) Phylogenetic analysis of the plastid inverted repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *Int J Plant Sci* 172: 541–558.
- Soltis DE, Smith S, Cellinese N, Refulio-Rodriguez NF, Olmstead R, et al. (2011) Inferring angiosperm phylogeny: a 17-gene analysis. *Am J Bot* 98: 704–730.
- Qiu YL, Li LB, Wang B, Chen Z, Knoop V et al. (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc Natl Acad Sci USA* 103: 15511–15516.
- Barniske AM, Borsch T, Müller K, Krug M, Worberg A, et al. (2012) Phylogenetics of early branching eudicots: Comparing phylogenetic signal across plastid introns, spacers, and genes. *J Syst Evol* 50: 85–108.
- Hoot SB, Magallón S, Crane PR (1999) Phylogeny of basal eudicots based on three molecular data sets: *atpB*, *rbcl* and 18S nuclear ribosomal DNA sequences. *Ann Mo Bot Gard* 86: 1–32.
- Soltis DE, Soltis PS, Chase MW, Mort M, Albach D, et al. (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcl* AND *atpB* sequences. *Bot J Linn Soc* 133: 381–461.
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA* 104: 19363–19368.
- Jansen RK, Cai Z, Raubeson LA, Daniell H, DePamphilis CW, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104: 19369–19374.
- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84: 9054–9058.
- Downie SR, Palmer JD (1992) Use of chloroplast DNA rearrangements in reconstructing plant phylogeny. In: Soltis PS, Soltis DE, Doyle JJ, eds. *Molecular Systematics of Plants*. New York: Chapman and Hall. 14–35.
- Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry R, ed. *Diversity and Evolution of Plants-genotypic Variation in Higher Plants*. Oxfordshire: CABI Publishing. 45–68.
- Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, et al. (2006) Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biol* 6: 17.
- Parks M, Cronm R, Liston A (2009) Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* 7: 84.



40. Diekmann K, Hodkinson TR, Wolfe KH, van den Bekerom R, Dix PJ, et al. (2009) Complete chloroplast genome sequence of a major allogamous forage species, perennial ryegrass (*Lolium perenne* L.). *DNA Res* 16: 165–76.
41. Kumar S, Hahn FM, McMahan CM, Cornish K, Whalen MC (2009) Comparative analysis of the complete sequence of the plastid genome of *Parthenium argentatum* and identification of DNA barcodes to differentiate *Parthenium* species and lines. *BMC Plant Biol* 9: 131.
42. Wu F-H, Chan M-T, Liao D-C, Hsu C-T, Lee Y-W, et al. (2010) Complete chloroplast genome of *Oncidium Gower* Ramsey and evaluation of molecular markers for identification and breeding in *Oncidiaceae*. *BMC Plant Biol* 10: 68.
43. Whittall JB, Syring J, Parks M, Buenrostro J, Dick C, et al. (2010) Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Mol Ecol* 19: 100–114.
44. Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26: 1135–1145.
45. Stull GW, Moore MJ, Mandala VS, Douglas N, Kates H-R, et al. (2013) A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. *App Plant Sci*: in press.
46. Zhang YJ, Ma PF, Li DZ (2011) High-Throughput Sequencing of Six Bamboo Chloroplast Genomes: Phylogenetic Implications for Temperate Woody Bamboos (Poaceae: Bambusoideae). *PLoS ONE* 6: e20596.
47. Steele PR, Hertweck KL, Mayfield D, McKain MR, Leebens-Mack J, et al. (2012) Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *Am J Bot* 99: 330–348.
48. Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, et al. (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am J Bot* 99: 349–364.
49. McCauley DE, Stevens JE, Peroni PA, Raveill JA (1996) The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *Am J Bot* 83: 727–31.
50. Small RL, Cronn RC, Wendel JF (2004) Use of nuclear genes for phylogeny reconstruction in plants. *Aust Syst Bot* 17: 145–70.
51. Provan J, Powell W, Hollingsworth PM (2001) Chloroplast microsatellites: new tools for studies in plant ecology and evolution. *Trends Ecol Evol* 16: 142–147.
52. Kim YK, Park CW, Kim KJ (2009) Complete chloroplast DNA sequence from a Korean endemic genus, *Megaleranthus sanciculifolia*, and its evolutionary implications. *Mol Cells* 27: 365–381.
53. Shinozaki K, Ohem M, Tanaka M, Wakasugi T, Hayashida N, et al. (1986) The complete nucleotide sequence of tobacco chloroplast genome: its gene organization and expression. *EMBO Journal* 5: 2043–2049.
54. Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Vasil IK, Bogorad L, eds. *Cell Culture and Somatic Cell Genetics in Plants*, Vol. 7A, The Molecular Biology of Plastids. San Diego, USA: Academic Press. 5–53.
55. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, et al. (2006) The complete chloroplast genome sequence of *Pelargonium x hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol Biol Evol* 23: 2175–2190.
56. Cosner ME, Jansen RK, Palmer JD, Downie SR (1997) The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. *Curr Genet* 31: 419–429.
57. Knox EB, Palmer JD (1999) The chloroplast genome arrangement of *Lobelia thuliniana* (Lobeliaceae): Expansion of the inverted repeat in an ancestor of the Campanulales. *Plant Syst Evol* 214: 49–64.
58. Plunkett GM, Downie SR (2000) Expansion and contraction of the chloroplast inverted repeat in *Apiaceae* subfamily *Apioideae*. *Syst Bot* 25: 648–667.
59. Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, et al. (2008) Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J Mol Evol* 66: 555–564.
60. Ozaki K (1987) *Tetracentron* leaves from the Neogene of Japan. *Transactions and Proceedings of the Palaeontological Society, Japan*, NS 146: 77–87.
61. Suzuki M, Joshi L, Noshira S (1991) *Tetracentron* wood from the Miocene of Noto Peninsula, Central Japan, with a short revision of homoxyllic fossil woods. *Bot Mag Tokyo* 104: 34–48.
62. Manchester SR, Chen I (2006) *Tetracentron* fruits from the Miocene of western North America. *Int J Plant Sci* 167: 601–605.
63. Pigg KB, Dillhoff RM, DeVore ML, Wehr WC (2007) New diversity among the Trochodendraceae from the Early/Middle Eocene Okanogan Highlands of British Columbia, Canada, and northeastern Washington State, United States. *Int J Plant Sci* 168: 521–532.
64. Chelebaeva AI, Chigayeva GB (1988) The genus *Trochodendron* (Trochodendraceae) in Miocene of Kamchatka. *Bot Zh* 73: 315–318.
65. Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperm re-revisited. *Am J Bot* 97: 1296–1303.
66. Anderson CL, Bremer K, Friis EM (2005) Dating phylogenetically basal eudicots using *rbcl* sequences and multiple fossil reference points. *Am J Bot* 92: 1737–1748.
67. Powell W, Morgante M, Mcdevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proc Natl Acad Sci USA* 92: 7759–7763.
68. Grassi F, Labra M, Scienza A, Imazio S (2002) Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. *Vitis* 41: 157–158.
69. Ebert D, Peakall R (2009) Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. *Mol Ecol Res* 9: 673–690.
70. Yang Z, Lu R, Tao C, Chen S, Ji Y (2012) Microsatellites for *Tetracentron sinense* (Trochodendraceae), a Tertiary relict endemic to East Asia. *Am J Bot* 99: e320–e322.
71. Fu LG (1992) *China Plant Red Data Book* (Vol.1). Beijing: Science Press.
72. Li R, Zhu H, Ruan J, Qian W, Fang X, et al. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20: 265–272.
73. Wyman SK, Jansen RK, Boore JL (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.
74. Conant GC, Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* 24: 861–862.
75. Rozas J, Sánchez-Delbarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
76. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273–W279.
77. Thurston MI, Field D (2005) Msatfinder: detection and characterisation of microsatellites, version 1.6.8.
78. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9: 772.
79. Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.
80. Felsenstein J (1985) Confidence limits on phylogeny: An approach using the bootstrap. *Evolution* 39: 783–791.
81. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29: 1969–1973.
82. Doyle JA (1992) Revised palynological correlations of the lower Potomac Group (USA) and the Cocobeach sequence of Gabon (Barremian-Aptian). *Cretac Res* 13: 337–349.
83. Hughes NF (1994) *The Enigma of Angiosperm Origins*. Cambridge, UK: Cambridge Univ Press.
84. Brenner GJ (1996) *Flowering Plant Origin, Evolution and Phylogeny*. New York: Chapman and Hall. 91–115.
85. Friis EM, Pedersen KR, Crane PR (1999) Early angiosperm diversification: The diversity of pollen associated with angiosperm reproductive structures in Early Cretaceous floras from Portugal. *Ann Mo Bot Gard* 86: 259–296.
86. Doyle JA, Hotton CL (1991) Pollen and Spores: Patterns of Diversification. Oxford: Clarendon. 169–195.
87. Magallón S, Castillo A (2009) Angiosperm diversification through time. *Am J Bot* 96: 349–365.
88. Friis EM, Eklund H, Pedersen KR, Crane PR (1994) *Virginianthus calycanthoides* gen. et sp. nov. – A calycanthaceous flower from the Potomac Group (Early Cretaceous) of eastern North America. *Int J Plant Sci* 155: 772–785.
89. Crane PR, Pedersen KR, Friis EM, Drinnan AN (1993) Early Cretaceous (early to middle Albian) platanoid infl. orescences associated with *Sapindopsis* leaves from the Potomac Group of North America. *Syst Bot* 18: 328–344.