



Article

Whole-Genome Comparison Reveals Heterogeneous Divergence and Mutation Hotspots in Chloroplast Genome of *Eucommia ulmoides* Oliver

Wencai Wang ¹, Siyun Chen ² and Xianzhi Zhang ^{3,*}

¹ Institute of Clinical Pharmacology, Guangzhou University of Chinese Medicine, Guangzhou 510000, China; wencaiwang@gzucm.edu.cn

² Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China; chensiyun@mail.kib.ac.cn

³ College of Forestry, Northwest A&F University, Yangling 712100, China

* Correspondence: zhangxianzhi@nwsuaf.edu.cn; Tel.: +86-29-8708-2230

Received: 1 March 2018; Accepted: 25 March 2018; Published: 30 March 2018



Abstract: *Eucommia ulmoides* (*E. ulmoides*), the sole species of Eucommiaceae with high importance of medicinal and industrial values, is a Tertiary relic plant that is endemic to China. However, the population genetics study of *E. ulmoides* lags far behind largely due to the scarcity of genomic data. In this study, one complete chloroplast (cp) genome of *E. ulmoides* was generated via the genome skimming approach and compared to another available *E. ulmoides* cp genome comprehensively at the genome scale. We found that the structure of the cp genome in *E. ulmoides* was highly consistent with genome size variation which might result from DNA repeat variations in the two *E. ulmoides* cp genomes. Heterogeneous sequence divergence patterns were revealed in different regions of the *E. ulmoides* cp genomes, with most (59 out of 75) of the detected SNPs (single nucleotide polymorphisms) located in the gene regions, whereas most (50 out of 80) of the indels (insertions/deletions) were distributed in the intergenic spacers. In addition, we also found that all the 40 putative coding-region-located SNPs were synonymous mutations. A total of 71 polymorphic cpDNA fragments were further identified, among which 20 loci were selected as potential molecular markers for subsequent population genetics studies of *E. ulmoides*. Moreover, eight polymorphic cpSSR loci were also developed. The sister relationship between *E. ulmoides* and *Aucuba japonica* in Garryales was also confirmed based on the cp phylogenomic analyses. Overall, this study will shed new light on the conservation genomics of this endangered plant in the future.

Keywords: *Eucommia ulmoides*; chloroplast genome; heterogeneous divergence; mutation hotspots; whole-genome comparison

1. Introduction

There are profuse paleoendemics (e.g., Eucommiaceae) and/or phylogenetically primitive taxa (e.g., Cercidiphyllaceae) in China due to the glaciation refuge role played during the Quaternary period [1,2]. Unfortunately, up to *circa* 5000 flora species are currently endangered in China, some of which have already become extinct [3]. Many plant species with important medicinal values have also been threatened seriously due to the increasing demand for raw materials of medicines, over-harvesting and habitat-loss [4–6]. Conservation of medicinal plants has become one of the most urgent issues faced today in China.

Eucommia ulmoides Oliver, a dioecious woody plant endemic to China, is the sole species in the family Eucommiaceae [7]. *E. ulmoides* has been widely cultivated and used as a herbal drug to reduce blood pressure and strengthen the body in central and southern China for at least 2000 years [8,9]. *E. ulmoides*

is also well-known as a “hardy rubber” tree that produces trans-polyisoprene rubber (i.e., gutta or Eu-rubber) in the leaves, bark, and pericarp [10,11]. It has been shown that *Eucommia* fossils occurred widely across the Northern Hemisphere from the Palaeocene onwards [12], which indicates that *E. ulmoides* is a representative model of Tertiary relict species, i.e., living from Tertiary to present. However, *E. ulmoides* may have been extinct in the wild and already listed in the Red List of Endangered Plant Species in China probably due to exhaustive human exploration [13,14]. Therefore, effective strategies are urgently needed to conserve this rare and endangered medicinal plant.

To date, studies on *E. ulmoides* have mainly focused on the morphological variation and the natural products [8,15]. Molecular and population genetics studies of this valuable tree lag behind largely due to limited DNA sequence resources [16,17]. Recently, nuclear microsatellites (nrSSR) were developed to investigate the genetic diversity of *E. ulmoides* [18,19]. Amplified fragment length polymorphism (AFLP) and sequence-related amplified polymorphism (SRAP) have been used to construct genetic maps of *E. ulmoides* [20,21]. The genetic markers of random amplified polymorphic DNA (RAPD), chloroplast microsatellite (cpSSR) and inter-simple sequence repeat (ISSR) have also been uncovered [22–24]. Nevertheless, the variability of these developed fingerprinting markers in *E. ulmoides* is relatively low, with limited population genetics information. A new and promising marker type i.e., Single Nucleotide Polymorphism (SNP) has gained high popularity during the last two decades [25,26]. With the on-going progress of high throughput sequencing techniques, it has become convenient to collect large-scale SNP data for genetic analyses [27,28]. Using SNP markers in conservation genetics studies of endangered plants has attracted much attention; for instance, in *Pinus ponderosa* Douglas ex Lawson [29], and *Sciadopitys verticillata* (Thunb.) Siebold and Zucc [30].

Chloroplast (cp) DNA sequences have been extensively used in the studies of plant population genetics and molecular phylogenetics [31–33]. Typically, cp genomes of land plants have a quadripartite structure with a pair of inverted repeats (IRs) separating a large single-copy (LSC) region and a small single-copy (SSC) region, ranging from 115 to 165 kilobase (kb) [34]. The cp genomes in general are inherited uniparentally, mostly maternally and are essentially recombination-free, leading to a smaller effective population size and a shorter coalescent time than the nuclear genomes [35]. Recently Wang et al. [17] reported a cp genome sequence of *E. ulmoides* with a length of 163,341 bp. Clearly, the availability of additional sequenced cp genomes from *E. ulmoides* would aid our understanding of the cp genome-wide variation at the individual level. Through comparative genomic analysis, polymorphic cpDNA loci with plentiful SNPs and indels i.e., nucleotide insertions and deletions can also be detected, which would be useful for further population genetics studies of *E. ulmoides*.

Genome skimming is currently one of the most economical techniques to obtain plastome sequences [36], through which obtaining complete cp genomes for plant phylogenomics inference becomes convenient [37]. In this study, we generated and characterized one complete cp genome of *E. ulmoides* using the genome skimming approach. By comparing the cp genome generated in this study and the one published previously [17], our main goals were to: (1) test whether the cp genomes in *E. ulmoides* show structural rearrangements; (2) reveal the divergence pattern of the cp genome in *E. ulmoides*; (3) identify highly variable cp genome-wide markers for subsequent population genetics studies of *E. ulmoides*.

2. Results

2.1. Chloroplast Genome Variation in *E. ulmoides*

About 20 million clean reads (4.72 Gb data) were generated from genome skimming sequencing. Two assembly methods (CLC Genomics Workbench and SPAdes software) both obtained the complete cp genome of *E. ulmoides* with high genome coverage (>180×) and there is no difference between the two assembled sequences, suggesting a high-quality cp genome map was achieved. The final cp genome size was determined to be 163,586 bp, similar to the previously published one (KU204775) (Table 1). The number of protein-coding genes, tRNA genes and rRNA genes, were the same as those in

the available *E. ulmoides* cp genome (KU204775). We have deposited the newly sequenced *E. ulmoides* cp genome in GenBank with accession number MF766010. The genome skimming sequencing reads have also been deposited in the Sequence Read Archive (SRA) with the accession number PRJNA399774.

Table 1. Comparison between the newly and previously sequenced chloroplast genomes of *Eucommia ulmoides*.

Item	This Study	KU204775
Chloroplast genome size (bp)	163,586	163,341
LSC ^a length (bp)	86,764	86,592
SSC ^b length (bp)	14,166	14,149
IRa/IRb ^c length (bp)	31,328	31,300
Number of genes (unique genes)	136 (115)	136 (115)
Number of protein-coding genes (unique genes)	89 (80)	89 (80)
Number of tRNA genes (unique genes)	39 (31)	39 (31)
Number of rRNA genes (unique genes)	8 (4)	8 (4)
GC ^d content (%)	38.33%	38.34%
Protein-coding regions (%)	51.91%	51.99%

^a LSC, large single-copy region; ^b SSC, small single-copy region; ^c IRa/IRb, two identical inverted repeat regions a/b; ^d GC, Guanine and Cytosine.

The whole genome alignments from MAFFT (Figure 1A) and MAUVE (Figure 1B) were consistent. There were no large genome rearrangements in the two cp genomes, which indicated that the cp genome structure in *E. ulmoides* is highly conserved and perfectly syntenic (Figure 1). Interestingly, small-scale nucleotide insertions and deletions were detected in the *E. ulmoides* cp genome. We found 15 insertions with more than ten nucleotides (11–111 bp) in the two cp genomes (Table 2). Five deletions in the range of 16–90 bp were also uncovered. It is worth noting that all of these insertions and deletions were involved in repeat sequence expansions and contractions (Table 2). Furthermore, across the entire cp genome of *E. ulmoides*, the sequence divergences were not uniform but highly heterogeneous (Figure 1).

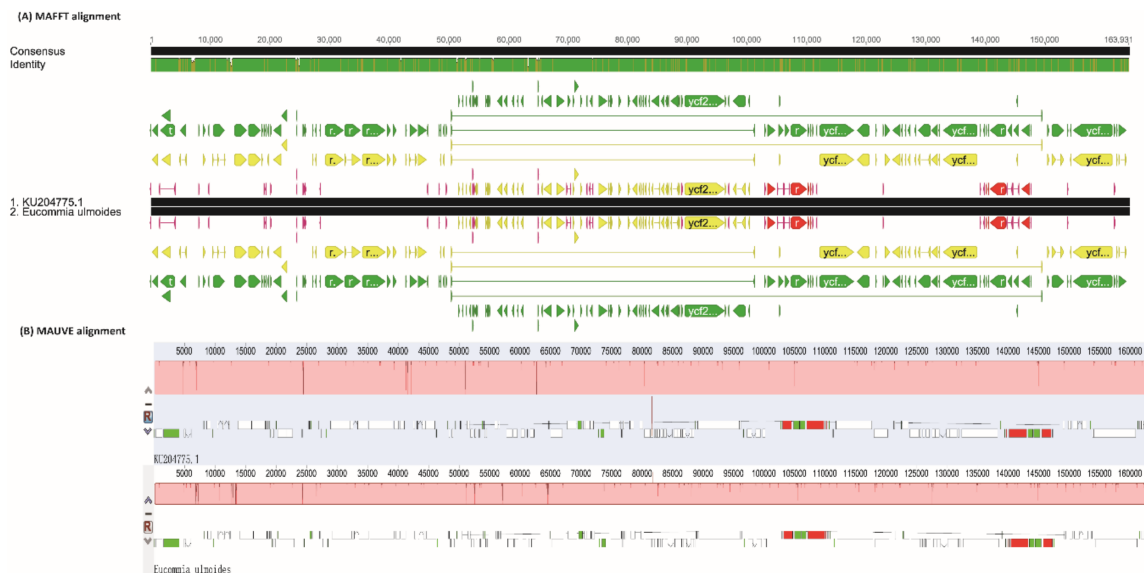


Table 2. DNA insertions and deletions with more than 10 nucleotides in the chloroplast genomes of *Eucommia ulmoides*.

No.	Size (bp)	Start Position	Location	Type
1	56	6851	<i>rps16-trnT(UGU)</i>	insertion
2	27	7006	<i>rps16-trnT(UGU)</i>	insertion
3	45	7196	<i>rps16-trnT(UGU)</i>	insertion
4	13	12,693	<i>ycf3-psaA</i>	insertion
5	23	12,912	<i>ycf3-psaA</i>	insertion
6	111	13,312	<i>ycf3-psaA</i>	insertion
7	12	13,471	<i>ycf3-psaA</i>	insertion
8	32	24,279	<i>psbD-trnT(GGU)</i>	insertion
9	12	26,615	<i>trnD(GUC)-psbM</i>	insertion
10	11	51,194	<i>rps12-rpl20</i>	insertion
11	17	52,547	<i>rps18-rpl33</i>	insertion
12	40	57,075	<i>psbJ-petA</i>	insertion
13	18	64,506	<i>accD-trnM(CAU)</i>	insertion
14	12	73,717	<i>trnL(UAA) intron</i>	insertion
15	14	127,486	<i>ndhG-ndhI</i>	insertion
16	16	4673	<i>trnK(UUU)-rps16</i>	deletion
17	44	24,700	<i>trnG(TTU)-trnE(UUC)</i>	deletion
18	31	41,767	<i>atpI-atpH</i>	deletion
19	44	51,109	<i>rps12-rpl20</i>	deletion
20	90	62,865	<i>accD-trnM(CAU)</i>	deletion

2.2. Molecular Marker Development

A total of 155 mutational events, including 75 nucleotide substitutions (SNPs) and 80 nucleotide indels (insertions and deletions), were detected within 71 loci of the cp genome in *E. ulmoides* (Figure 2). There were 98 mutations (51 SNPs and 47 indels) and 15 mutations (12 SNPs and 3 indels) in the LSC and SSC regions, respectively. In addition, 42 mutations (12 SNPs and 30 indels) were located in the IR region. Distribution patterns of SNPs and indels differed largely in the cp genic and intergenic regions of *E. ulmoides*. Most of the SNPs (59 out of 75) were found in the gene sequences, including 31 protein-coding genes and one tRNA gene. In contrast, indels were mainly (50 out of 80) distributed in the intergenic spacers (Figure 2). Upon further investigation of SNPs and indels in the nine intron-containing protein-coding genes (*atpF*, *ndhA*, *ndhB*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *rps16*, *ycf3*), we found that all the mutations were located in the intron regions. In all, 40 SNPs and 14 indels occurred in the plastid-coding sequence (CDS) regions.

The proportion of variability in the 71 polymorphic loci ranged from 0.03% to 1.55% with a mean value of 0.37% (Figure 3). The mutation rates in most (53 out of 71) of the loci were between 0.10 and 1.00%. Five of these DNA fragments i.e., *atpF-atpA*, *rps18-rpl33*, *psaJ*, *infA* and *rpl32* had variations exceeding 1.00%. Considering the relatively high percentage of variability and convenience for primer design in PCR (Polymerase Chain Reaction) and sequencing experiments, we chose 20 highly variable loci with length of 200–1500 bp as potential molecular markers for subsequent population genetic studies (Table 3). The percentage of variations in these 20 loci all exceeded 0.25%, among which 16 had a percentage of variable characters (VCs) greater than 0.30% (Table 3).

Through SSR analysis, we found a total of 31 SSR loci in the newly assembled *E. ulmoides* cp genome, among which 27 were shared by the two genomes. Further detection revealed that eight cpSSR loci were polymorphic in *E. ulmoides* (Table 4). All the polymorphic cpSSR loci were mononucleotide repeats, ranging from 10–15 bp in length. Five polymorphic cpSSR loci were located in the LSC region, with another three ones in the IR regions (Table 4).

Table 3. The 20 chloroplast DNA fragments with relative high genetic divergences identified in *Eucommia ulmoides*.

Region	Aligned Length (bp)	No. VCs ^a	Percentage of VCs (%)
<i>infA</i>	234	3	1.28
<i>rps18-rpl33</i>	343	4	1.17
<i>rps12</i>	369	3	0.81
<i>rrn5-trnR(ACG)</i>	266	2	0.75
<i>ycf15</i>	210	1	0.48
<i>trnI(GAU)</i>	1062	5	0.47
<i>petB</i>	651	3	0.46
<i>ycf3-psaA</i>	1502	6	0.40
<i>trnT(GGU)-atpE</i>	261	1	0.38
<i>ndhG</i>	531	2	0.38
<i>ycf4</i>	546	2	0.37
<i>trnG(UCC)-psbZ</i>	280	1	0.36
<i>rps16</i>	1170	4	0.34
<i>trnA(UGG)</i>	881	3	0.34
<i>ndhB-rps7</i>	325	1	0.31
<i>psbK-trnQ(UUG)</i>	338	1	0.30
<i>trnI(CAU)-ycf2</i>	357	1	0.28
<i>ycf15-trnL(CAA)</i>	359	1	0.28
<i>psbB</i>	1521	4	0.27
<i>rpl20</i>	384	1	0.26

^a VCs: variable characters, including SNPs and indels.

Table 4. The polymorphic chloroplast SSRs identified in *Eucommia ulmoides*.

No.	SSR Repeat Motif	Length Variation (bp)	Location	Region ^a
1	(G)	10–11	<i>trnG(UCC)-psbZ</i>	LSC
2	(A)	12–15	<i>rpoC2</i>	LSC
3	(A)	12–13	<i>ycf1</i>	IRb
4	(A)	13–14	<i>rpl32-trnL(UAG)</i>	IRa
5	(T)	10–14	<i>psbJ-petA</i>	LSC
6	(T)	10–11	<i>trnG(GCC)-trnS(GCU)</i>	LSC
7	(T)	12–13	<i>ycf1</i>	IRa
8	(T)	14–15	<i>rpl16-rps3</i>	LSC

^a LSC, large single-copy region; IRa/IRb, two identical inverted repeat regions a/b.

2.3. SNP Calling and Phylogenomic Inference

The SNPs calling analysis using the previously published *E. ulmoides* cp genome (KU204775) as reference revealed a total of 75 SNPs. This result of SNP occurrence was consistent with the aforementioned molecular marker analysis (75 SNPs, Figure 2), which indicated that the detected SNPs were really present in different individuals of *E. ulmoides*. Further examination of the 40 SNPs in the CDS regions suggested that all these SNPs were synonymous, i.e., no amino acid change at the protein level. There were 34 transitions and six transversions in the protein-coding region SNPs. The average frequency of SNPs occurrence in the *E. ulmoides* cp genome was calculated as 0.46 per kb.

The final length of the supermatrix dataset contained 96,894 unambiguously aligned nucleotide characters. Three methods produced a congruent phylogenetic tree, shown in Figure 4. Eudicots, monocots and magnoliids were all highly supported to be monophyletic (94–100/90–100/1.00). Magnoliids diverged firstly, followed by monocots, then Chloranthales and eudicots, with relatively low (60–72/74–80/0.90–0.99) support values. Two *E. ulmoides* individuals clustered together with high statistical support values (100/100/1.00), which was subsequently sister to *Aucuba japonica* Thunb.

in the Garryales clade with high statistical support values (100/100/1.00). Garryales, Gentianales and Solanales formed as the highly supported lamiids lineage (100/100/1.00), resolved as (Garryales, (Gentianales, Solanales)). Asterales and Dipsacales formed a highly supported clade, campanulids (100/100/1.00), which was sister to lamiids (100/100/1.00) in asterids. Ericales was resolved as the basal most group in the asterids lineage (100/100/1.00). Brassicales, Malvales, Myrtales and Sapindales clustered in the malvids clade (95/98/1.00) with resolution as (((Brassicales, Malvales), Sapindales), Myrtales). Fabales, Malpighiales and Rosales were located in the fabids lineage (98/100/1.00), which was a sister clade to malvids in rosids.

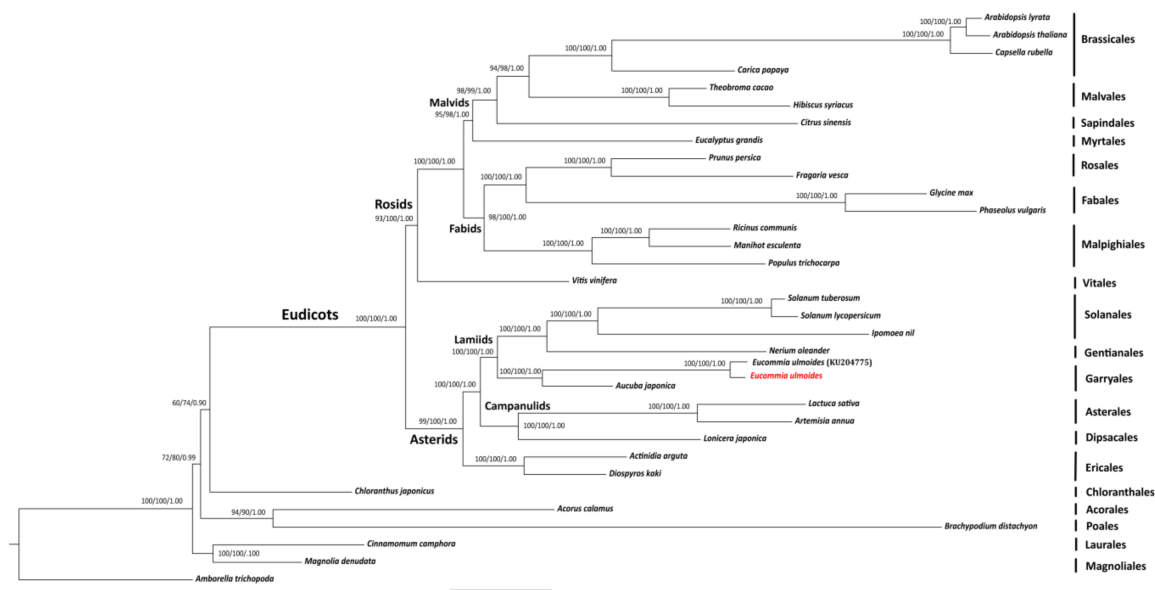


Figure 4. Maximum likelihood (ML) tree for 34 taxa based on 80 unique plastid protein-coding genes of *Eucommia ulmoides*. Values above the branches represent maximum parsimony bootstrap (MPBS)/maximum likelihood bootstrap (MLBS)/Bayesian inference posterior probability (PP). The newly sequenced *Eucommia ulmoides* chloroplast genome is indicated by red color and the previously published *E. ulmoides* chloroplast genome is followed by its GenBank accession number KU204775.

3. Discussion

3.1. Conserved Chloroplast Genome Structure in *E. ulmoides*

Land plant cp genomes are generally inherited as a haplotype with no recombination, providing useful genetic information to trace relationships between different species [35,38]. Within species cp genome structure was highly conserved [39]. As expected, it is the case in *E. ulmoides* in terms of the contained genes and coding regions in the cp genomes (Table 1). Further whole-genome alignments suggested that the two cp genomes of *E. ulmoides* did not show genome rearrangement having the same linear gene order (Figure 1). As such it is reasonable to use cp genomes for subsequent conservation genomics studies on *E. ulmoides*.

It is noteworthy that the newly sequenced cp genome of *E. ulmoides* (163,586 bp) in this study is 245 bp larger than that of the previously reported one (163,341 bp, [17]) (Table 1). The cp genome size variation within different individuals of the same species has been reported for several other plants, such as in *Camptotheca acuminata* Decne. with its size varied as 157,806 bp [40], 157,877 bp [41] and 162,382 bp [42]. Nuclear genome size variations in plants are mostly caused by the repeats activities (e.g., expansions/contractions) via illegitimate recombination in addition to polyploidy [43–47]. In the two *E. ulmoides* cp genomes, we detected 15 insertions and five deletions, with more than 10 nucleotides for each (Table 2). All these sequences were observed to be part of or the whole DNA

repeats. For instance, the repeat sequences in *rps16-trnT(UGU)*, *psbD-trnT(GGU)*, *rps12-rpl20* and *accD-trnM(CAU)* have been detected in the study of Wang et al. [17] as well. Therefore, potentially the illegitimate recombination between repeat regions of *E. ulmoides* cp genome may contribute to the cp genome size variation.

3.2. Heterogeneous Divergence in *E. ulmoides* Chloroplast Genome

Heterogeneous divergence patterns in cp genomes have been reported in several plant groups, such as in Actinidiaceae [37] and in Poaceae [48]. The alignment of the two available cp genomes of *E. ulmoides* revealed highly heterogeneous sequence divergences within this species (Figures 1 and 2). All the identified SNPs and indels from the intron-containing protein-coding genes were located in the intron regions. Due to natural selection CDS regions are in general more conserved than non-coding regions (i.e., intergenic sequences and introns) [49]. In addition, nucleotide substitutions likely have less destructive effect to the integrity of open reading frame (ORF) than indels [50]. We, thus, speculated that this functional constraint may lead to the contrasting occurrences of SNPs and indels in the *E. ulmoides* cp genomes.

The occurrence of synonymous SNPs was more abundant than that of the non-synonymous SNPs in CDS regions because of selection process [51,52]. As expected, all the SNPs detected in protein-coding regions were synonymous. Moreover, since the transitions rather than the transversions usually would generate more synonymous mutations in the CDS the transition SNPs are more easily retained than the transversion ones [52,53]. In this study we found that the transition SNPs (34) were indeed more frequently detected than the transversion SNPs (6). Previous studies have reported a high level of nuclear genetic diversity at the population level of *E. ulmoides* [18,54]. In this study, we revealed that the frequency of plastid SNPs were 0.46 per kb at the whole cp genome level, lower than the average of 1.02 per kb in the nuclear genes of *E. ulmoides* [54]. The difference of SNP frequency between the cp genome and the nuclear genes could be caused by insufficient sampling in this study and/or different variation rates between the plastid and nuclear sequences.

3.3. Mutation Hotspots in *E. ulmoides* Chloroplast Genome

In general, protein-coding genes in the cp genome have lower sequence variation than the non-coding loci, for instance in bamboos [55] and mimosoid legume [56]. However, an accelerated variation rate of some plastid protein-coding genes has been reported, such as the *psb* in Poaceae [51], *rps* in Saxifragales [57], and *accD* and *rpl20* in Actinidiaceae [37]. In *E. ulmoides*, we found three protein-coding genes (*infA*, *psaJ* and *rpl32*) that varied the most quickly, all having variations exceeding 1.2% (Figure 3). The gene *infA* encodes translation initiation factor 1 and has been found missing in cp genomes of several plant lineages, e.g., in rosids [58]. The other two genes i.e., *psaJ* and *rpl32* code for photosystem I protein J and ribosomal protein L32, respectively, both of which are short in length with the former having 129 bp and the latter 138 bp (this study and [17]). The relatively high level of variation of these genes in *E. ulmoides* indicates that they are less constrained. Abnormal DNA replication, repair or recombination [59,60] may lead to the elevated divergence of these genes.

The genetic markers of SSR, AFLP and SRAP have been developed and used for the population genetics studies of *E. ulmoides* [18,20,21]. However, the above fingerprinting markers [23,24] may provide insufficient genetic information to resolve the population structure and history of *E. ulmoides*. The relationships among natural and cultivated populations of *E. ulmoides* are elusive at present [18,24]. SNP markers are ample in plant cp genomes, making them useful candidates for population genetics studies. Using cp genome SNP markers has widely received attention during the past few years with the advances of high throughput sequencing techniques [61,62]. Highly variable cpDNA fragments have been mined for phylogenetic and population genetic studies in several species using cp genomes data, such as in kiwifruit [37] and temperate woody bamboos [55]. Given that it would be easy to amplify and sequence DNA fragments with length from *circa* 200 to 1500 bp using Sanger sequencing method [62,63], we thus chose 20 cpDNA loci with relatively high genetic divergences as potential

molecular markers (Table 3) for subsequent population genomics studies of *E. ulmoides*. These selected plastid genome-wide loci would genetically be informative for uncovering the genetic relationships among the natural and cultivated *E. ulmoides* populations.

Additionally, 27 cpSSR loci identified by Wang et al. [17] were also confirmed in our SSR analysis, among which eight were further mined as polymorphic cpSSR loci (Table 4). Given that polymorphic cpSSR loci could be applied as useful markers to meet certain study purposes under the circumstances of limited budget [64,65], the newly developed polymorphic cpSSR loci in *E. ulmoides* here would be potential genetic markers to facilitate subsequent population genetics studies in the future.

3.4. Phylogenomic Validation of *E. ulmoides*

The newly obtained *E. ulmoides* cp genome was further validated via phylogenomic analyses using 34 complete plastomes from 10 major lineages of angiosperms. The resulting phylogenomic tree highly supported the clade of two *E. ulmoides* cp genomes (Figure 4), confirming the validity of the assembled and annotated cp genome of *E. ulmoides* in this study. The sister relationship between *E. ulmoides* and *A. japonica* in the Garryales clade was highly supported, which is consistent with the results derived from five organellar genes [66] and 36 plastid genes [17], supporting the classification of *E. ulmoides* (Eucommiaceae) in the updated APG IV system [67]. *E. ulmoides* and *A. japonica* are both woody and have unisexual flowers in separate individuals, which seem to be morphological synapomorphies for the order Garryales [68]. An average of 92.6% identities between 78 common unique cp protein-coding genes in *E. ulmoides* and *A. japonica* were also detected, suggesting a high similarity between the two species at the molecular level. Garryales was shown to be closely related to the clade of (Gentianales + Solanales) in lamiids, in line with the APG IV system [67].

All 20 sampled orders were highly supported to be monophyletic separately (Figure 4), agreeing with the APG IV system [67]. Within eudicots two large sister clades i.e., asterids and rosids were uncovered, and the relationships among these two lineages were highly resolved as ((campanulids, lamiids), Ericales) and ((fabids, malvids), Vitales), respectively as stated previously [69]. The branching patterns of species within campanulids and lamiids are consistent with recent studies [66,70] as (((Gentianales, Solanales), Garryales), (Asterales, Dipsacales)). Our analyses also resolved the phylogenetic relationships within fabids and malvids as (((Brassicales, Malvales), Sapindales), Myrtales), ((Fabales, Rosales), Malpighiales)), consistent with the results of previous studies [17,69]. It is noteworthy that the branching orders of magnoliids, monocots, Chloranthales and eudicots only obtained low-level support values here (Figure 4). Further studies with expanded taxon samples are expected to confirm the phylogeny of these lineages. Moreover, plastome is inherited uniparentally in general, which might introduce biases to species phylogeny inference [71,72]. Analyses using orthologous nuclear genes are also needed for studying the evolutionary history of *E. ulmoides* among the flowering plants [54].

4. Materials and Methods

4.1. Plant Materials and DNA Sequencing

Fresh healthy leaves were collected from an adult male individual of *E. ulmoides* growing in the Arboretum of Northwest Agriculture and Forest University in Yangling, Shanxi, China, in April 2015. After collection, the leaves were immediately immersed in liquid nitrogen and then stored at -80°C until use. The voucher specimen of this tree was deposited at the Trees Herbarium of Northwest A and F University with accession number ZXZ15027.

Total genomic DNA was extracted by the CTAB method [73]. Paired-end (PE) libraries with insert size *circa* 500 bp were constructed from fragmented genomic DNA based on standard Illumina protocols (Illumina Inc., San Diego, CA, USA). Prepared library was then sequenced for PE 100 bp read length on the Illumina HiSeq 2000 platform at the Beijing Genomics Institute (BGI) in Shenzhen, China.

4.2. Genome Assembly and Annotation

Fastq format PE reads were supplied with adaptor sequences removed. Poor quality reads with phred scores lower than 20 for more than 10% of their bases were also removed. Two independent methods were used to assemble the *E. ulmoides* cp genome. (1) The cp genome was de novo assembled using the CLC Genomics Workbench v7.5 software (CLC Bio, Aarhus, Denmark) based on the clean reads. After discarding contigs with length <300 bp and sequences with coverage <50, the remaining contigs were searched against the available cp genome of *E. ulmoides* (GenBank accession number KU204775) that used as the reference by BLAST (<http://blast.ncbi.nlm.nih.gov/>) with *e*-value <10⁻⁵. Aligned contigs with ≥90% similarity and query coverage were determined as cpDNA sequences and ordered according to the reference genome. Small gaps were filled using PE clean reads as conducted in Wang et al. [37]. (2) The clean reads were firstly mapped to the reference cp genome of *E. ulmoides* to determine the proportion of cpDNA using Bowtie v2.3.1 program [74] with a maximum of 3 mismatches. Subsequently, we applied SPAdes v3.9 software [75] with default setting to assemble the cp genome using the determined cpDNA clean reads.

DOGMA software [76] was used for initial cp genome annotation. Start/stop codons and intron/exon boundaries were checked and adjusted manually when necessary by comparing to the reference genome. tRNA genes were confirmed based on tRNAscan-SE 1.21 [77].

4.3. Genome-Wide Comparison and Divergent Hotspot Identification

The previously published cp genome of *E. ulmoides* (accession number: KU204775) was downloaded from GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>). This genome was aligned with the *E. ulmoides* cp genome described herein, using MAFFT program [78] and MAUVE software [79], respectively, and manually adjusted where necessary. The obtained pairwise alignment of the cp genomes was visualized in Geneious v9.0 [80]. Moreover, given the genome repeat sequences expansion and contraction may result in genome size variation [81], we examined the DNA insertions and deletions in repeat regions of the two *E. ulmoides* cp genomes.

The two *E. ulmoides* cp genomes were analyzed to identify molecular markers that can be selected in subsequent population genetic studies. We firstly extracted both the genic and intergenic DNA fragments in each cp genome using the “Extract Sequences” option in DOGMA [76]. Then the homologous loci were aligned individually by MUSCLE program (<http://www.drive5.com/muscle/>) [82] implemented in Geneious v9.0 [80] with default settings. Manual adjustments were made for the alignments where necessary. The proportion of mutational events for each genic and intergenic locus was calculated as follows: the proportion of variation = ((NS + ID)/L) × 100, where NS = the number of nucleotide substitutions (SNPs), ID = the number of indels (insertions and deletions), L = the aligned sequence length.

Polymorphic cpSSR loci were further mined by genome comparison. Firstly, SSRs in the newly sequenced *E. ulmoides* cp genome were detected by MISA perl script (<http://pgrc.ipk-gatersleben.de/misa/>). The parameter of minimum repeat unit was set as 10 for mono-, 6 for di-, and 5 for tri-, tetra-, penta-, and hexanucleotide SSRs [83]. Then, all the identified SSR loci were compared to the 29 cpSSR loci of Wang et al. [17] to develop polymorphic cpSSR markers in *E. ulmoides*.

4.4. SNPs Validation and Phylogenomic Analyses

To confirm the SNPs identified by the aforementioned cp genome alignment, we here mapped the genome skimming clean reads generated in this study to the previously published *E. ulmoides* cp genome (KU204775) [17] for SNP calling. Picard-tools v1.41 (<http://broadinstitute.github.io/picard/>) and samtools v0.1.18 [84] were applied to sort and remove duplicated reads and merge the bam alignment results. GATK3 software [85] was further used to perform SNPs identification. Raw vcf files were filtered with GATK standard filter method and other parameters were set as defaults. Moreover, to reveal if the coding region SNPs detected in *E. ulmoides* cp genome caused amino

acid substitution on protein level, we firstly translated each protein-coding gene into amino acids in Geneious v9.0 [80]. Then the protein sequences of each gene were aligned, respectively using MUSCLE [82]. The mutational events were checked to uncover the synonymous and nonsynonymous SNPs and the nucleotide transitions and transversions.

Phylogenomic analyses were also conducted to validate the newly assembled and annotated *E. ulmoides* cp genome. 34 plastomes representing 10 major lineages of angiosperms (Table S1) were included for phylogenomic analyses. *Amborella trichopoda* from basal angiosperm lineages was defined as outgroup according to previous studies [66,67]. 80 unique plastid protein-coding genes of *E. ulmoides* (Table S2) were used for the phylogenetic inferences. Each gene was aligned individually by MUSCLE [82] in Geneious v9.0 [80], and then concatenated as a supermatrix. Gaps were not included in the dataset.

Three methods i.e., maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) were used for phylogenetic reconstruction. We performed parsimony heuristic tree searches in PAUP v4.0b10 [86] with parameters set as 1000 random addition sequence replicates, tree bisection and reconnection (TBR) branch swapping, and MulTrees option in effect. 1000 bootstrap replicates [87] were calculated to evaluate the branch support (MPBS) of the MP tree. RAxML v.8.2.8 [88] and MrBayes 3.2.6 [89] in the CIPRES Science Gateway v3.3.3 [90] were applied for ML and BI analyses, respectively. Supermatrix was partitioned by genes and GTR + G model of nucleotide substitution was used. For the ML tree we conducted 1000 fast bootstrap ML reps to assess the support values (MLBS) of internal nodes. In Bayesian analysis two runs with four chains were carried out up to 50,000,000 generations, sampling one tree every 1000 generations till convergence, i.e., the average standard deviation of split frequencies <0.01. We discarded the first 25% of trees as burn-in, and used the remaining trees to estimate the majority-rule consensus BI tree and posterior probabilities (PP).

5. Conclusions

In summary, in the present study we generated one complete cp genome of *E. ulmoides* using the genome skimming approach. Through comprehensive genome-wide comparative analyses we found that the cp genomes within *E. ulmoides* were highly conserved in terms of structure and content. Nevertheless, obviously heterogeneous sequence divergences were revealed in different regions of the *E. ulmoides* cp genome. A total of 20 polymorphic DNA fragments and eight SSR loci have been identified as potential cpDNA markers for subsequent population genetics studies of this tree species. The phylogenetic placement of *E. ulmoides* in angiosperms was robustly resolved as well based on the cp genomes data, strongly supporting the sister relationship between *E. ulmoides* and *A. japonica* in the asterids lineage. The data presented here will aid further conservation genomic studies and facilitate the development of plastid genetic engineering for *E. ulmoides*.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1422-0067/19/4/1037/s1>. Table S1: Taxa and GenBank accession numbers included in the phylogenomic analyses, Table S2: List of 80 unique plastid protein-coding genes of *Eucommia ulmoides* included in the phylogenomic analyses.

Acknowledgments: We would like to thank Zhirong Zhang for the help in the experiment performance. This work is supported by the National Natural Science Foundation of China (31600173) and the Basic Science Fund of Northwest A&F University (2452016052).

Author Contributions: Wencai Wang and Xianzhi Zhang conceived and designed the experiments; Wencai Wang performed the experiments; Siyun Chen and Xianzhi Zhang analyzed the data; Xianzhi Zhang contributed reagents/materials/analysis tools; Wencai Wang and Xianzhi Zhang wrote the paper. All authors reviewed and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huang, H. Plant diversity and conservation in China: Planning a strategic bioresource for a sustainable future. *Bot. J. Linn. Soc.* **2011**, *166*, 282–300. [[CrossRef](#)] [[PubMed](#)]
2. Liu, J.; Ouyang, Z.; Pimm, S.L.; Raven, P.H.; Wang, X.; Miao, H.; Han, N. Protecting China's biodiversity. *Science* **2003**, *300*, 1240–1241. [[CrossRef](#)] [[PubMed](#)]
3. Lópezpujoll, J.; Zhang, F.M.; Ge, S. Plant biodiversity in China: Richly varied, endangered, and in need of conservation. *Biodivers. Conserv.* **2006**, *15*, 3983–4026. [[CrossRef](#)]
4. Gu, J. Conservation of plant diversity in China: Achievements, prospects and concerns. *Biol. Conserv.* **1998**, *85*, 321–327. [[CrossRef](#)]
5. Chen, S.L.; Hua, Y.; Luo, H.M.; Wu, Q.; Li, C.F.; Steinmetz, A. Conservation and sustainable use of medicinal plants: Problems, progress, and prospects. *Chin. Med.* **2016**, *11*, 37. [[CrossRef](#)] [[PubMed](#)]
6. Huang, L.; Yang, B.; Wang, M.; Fu, G. An approach to some problems on utilization of medicinal plant resource in China. *China J. Chin. Mater. Med.* **1999**, *24*, 70–73.
7. Zhang, Z.Y.; Zhang, H.D.; Turland, N.J. Eucommiaceae. In *Flora of China*; Wu, Z.Y., Raven, P.H., Hong, D.Y., Eds.; Science Press and Missouri Botanical Garden: Beijing, China, 2003; p. 43.
8. Kawasaki, T.; Uezono, K.; Nakazawa, Y. Antihypertensive mechanism of food for specified health use: "Eucommia leaf glycoside" and its clinical application. *J. Health Sci.* **2000**, *22*, 29–36.
9. Liu, H.; Hongyan, D.U.; Tana, W. Advances in research on biotechnology breeding of *Eucommia ulmoides*. *Hunan For. Sci. Technol.* **2016**, *43*, 132–136.
10. Suzuki, N.; Uefuji, H.; Nishikawa, T.; Mukai, Y.; Yamashita, A.; Hattori, M.; Ogasawara, N.; Bamba, T.; Fukusaki, E.; Kobayashi, A. Construction and analysis of EST libraries of the trans-polyisoprene producing plant, *Eucommia ulmoides* Oliver. *Planta* **2012**, *236*, 1405–1417. [[CrossRef](#)] [[PubMed](#)]
11. Du, H.Y.; Hu, W.Z.; Yu, R. *The Report on Development of China's Eucommia Rubber Resources and Industry (2014–2015)*; Social Sciences Academic Press: Beijing, China, 2015.
12. Manchester, S.R.; Chen, Z.D.; An-Ming, L.U.; Uemura, K. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *J. Syst. Evol.* **2009**, *47*, 1–42. [[CrossRef](#)]
13. Mabberley, D.J. *The Plant Book*; Cambridge University Press: Cambridge, UK, 1989.
14. Fu, L.G.; Jin, J.M. *Red List of Endangered Plants in China*; Science Press: Beijing, China, 1992.
15. Du, H.Y. *China Eucommia Pictorial*; China Forestry Publishing House: Beijing, China, 2014.
16. Wang, L.; Du, H.; Wuyun, T.N. Genome-wide identification of microRNAs and their targets in the leaves and fruits of *Eucommia ulmoides* using high-throughput sequencing. *Front. Plant Sci.* **2016**, *7*, 1632. [[CrossRef](#)] [[PubMed](#)]
17. Wang, L.; Wuyun, T.N.; Du, H.; Wang, D.; Cao, D. Complete chloroplast genome sequences of *Eucommia ulmoides*: Genome structure and evolution. *Tree Genet. Genomes* **2016**, *12*, 12. [[CrossRef](#)]
18. Zhang, J.; Xing, C.; Tian, H.; Yao, X. Microsatellite genetic variation in the Chinese endemic *Eucommia ulmoides* (Eucommiaceae): Implications for conservation. *Bot. J. Linn. Soc.* **2013**, *173*, 775–785. [[CrossRef](#)]
19. Zhang, W.R.; Li, Y.; Zhao, J.; Wu, C.H.; Ye, S.; Yuan, W.J. Isolation and characterization of microsatellite markers for *Eucommia ulmoides* (Eucommiaceae), an endangered tree, using next-generation sequencing. *Genet. Mol. Res.* **2016**, *15*. [[CrossRef](#)] [[PubMed](#)]
20. Li, Y.; Wang, D.; Li, Z.; Wei, J.; Jin, C.; Liu, M. A molecular genetic linkage map of *Eucommia ulmoides* and quantitative trait loci (QTL) analysis for growth traits. *Int. J. Mol. Sci.* **2014**, *15*, 2053–2074. [[CrossRef](#)] [[PubMed](#)]
21. Wang, D.; Li, Y.; Li, L.; Wei, Y.; Li, Z. The first genetic linkage map of *Eucommia ulmoides*. *J. Genet.* **2014**, *93*, 13–20. [[CrossRef](#)] [[PubMed](#)]
22. Wang, A.Q.; Huang, L.Q.; Shao, A.J.; Cui, G.H.; Chen, M.; Tong, C.H. Genetic diversity of *Eucommia ulmoides* by RAPD analysis. *China J. Chin. Mater. Med.* **2006**, *31*, 1583–1586.
23. Yao, X.; Deng, J.; Huang, H. Genetic diversity in *Eucommia ulmoides* (Eucommiaceae), an endangered traditional Chinese medicinal plant. *Conserv. Genet.* **2012**, *13*, 1499–1507. [[CrossRef](#)]
24. Yu, J.; Wang, Y.; Peng, L.; Ru, M.; Liang, Z.S. Genetic diversity and population structure of *Eucommia ulmoides* Oliver, an endangered medicinal plant in China. *Genet. Mol. Res.* **2015**, *14*, 2471–2483. [[CrossRef](#)] [[PubMed](#)]
25. Vignal, A.; Milan, D.; Sancristobal, M.; Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* **2002**, *34*, 275–305. [[CrossRef](#)] [[PubMed](#)]

26. Bernardi, J.; Mazza, R.; Caruso, P.; Reforgiato, R.G.; Marocco, A.; Licciardello, C. Use of an expressed sequence tag-based method for single nucleotide polymorphism identification and discrimination of *Citrus* species and cultivars. *Mol. Breed.* **2013**, *31*, 705–718. [[CrossRef](#)]
27. Elshire, R.J.; Glaubitz, J.C.; Sun, Q.; Poland, J.A.; Kawamoto, K.; Buckler, E.S.; Mitchell, S.E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **2011**, *6*, e19379. [[CrossRef](#)]
28. Kess, T.; Gross, J.; Harper, F.; Boulding, E.G. Low-cost ddRAD method of SNP discovery and genotyping applied to the periwinkle *Littorina saxatilis*. *J. Molluscan Stud.* **2016**, *82*, eyv042.
29. Potter, K.M.; Hipkins, V.D.; Mahalovich, M.F.; Means, R.E. Nuclear genetic variation across the range of ponderosa pine (*Pinus ponderosa*): Phylogeographic, taxonomic and conservation implications. *Tree Genet. Genomes* **2015**, *11*, 38. [[CrossRef](#)]
30. Worth, J.R.P.; Yokogawa, M.; Pérez-Figueroa, A.; Tsumura, Y.; Tomaru, N.; Janes, J.K.; Isagi, Y. Conflict in outcomes for conservation based on population genetic diversity and genetic divergence approaches: A case study in the Japanese relictual conifer *Sciadopitys verticillata* (Sciadopityaceae). *Conserv. Genet.* **2014**, *15*, 1243–1257. [[CrossRef](#)]
31. Chung, S.M.; Staub, J.E.; Lebeda, A.; Paris, H.S. Consensus chloroplast primer analysis: A molecular tool for evolutionary studies in Cucurbitaceae. In Proceedings of the Progress in Cucurbit Genetics and Breeding Research, Olomouc, Czech Republic, 12–17 July 2004.
32. Ahmed, I.; Matthews, P.J.; Biggs, P.J.; Naem, M.; Mclenachan, P.A.; Lockhart, P.J. Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa. *Mol. Ecol. Resour.* **2013**, *13*, 929–937. [[CrossRef](#)] [[PubMed](#)]
33. Zhang, Y.; Du, L.; Ao, L.; Chen, J.; Li, W.; Hu, W.; Wei, Z.; Kim, K.; Lee, S.C.; Yang, T.J. The complete chloroplast genome sequences of five *Epimedium* species: Lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* **2016**, *7*, 696. [[CrossRef](#)] [[PubMed](#)]
34. Raubeson, L.A.; Jansen, R.K. Chloroplast genomes of plants. In *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*; Henry, R.J., Ed.; CABI: Cambridge, MA, USA, 2005; pp. 45–68.
35. Birky, C.W., Jr. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. *Proc. Natl. Acad. Sci. USA* **1996**, *92*, 11331–11338. [[CrossRef](#)]
36. Straub, S.C.; Parks, M.; Weitemier, K.; Fishbein, M.; Cronn, R.C.; Liston, A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *Am. J. Bot.* **2012**, *99*, 349–364. [[CrossRef](#)] [[PubMed](#)]
37. Wang, W.C.; Chen, S.Y.; Zhang, X.Z. Chloroplast genome evolution in Actinidiaceae: *clpP* Loss, heterogenous divergence and phylogenomic practice. *PLoS ONE* **2016**, *11*, e0162324. [[CrossRef](#)] [[PubMed](#)]
38. Wicke, S.; Schneeweiss, G.M.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297. [[CrossRef](#)] [[PubMed](#)]
39. Wicke, S.; Schneeweiss, G.M. Next-Generation Organellar Genomics: Potentials and Pitfalls of High-Throughput Technologies for Molecular Evolutionary Studies and Plant Systematics. In *Next Generation Sequencing in Plant Systematics*; International Association for Plant Taxonomy (IAPT): Bratislava, Slovakia, 2015.
40. Chen, S.Y.; Zhang, X.Z. Characterization of the complete chloroplast genome of the relict Chinese false tupelo, *Camptotheca acuminata*. *Conserv. Genet. Resour.* **2017**, 1–4. [[CrossRef](#)]
41. Yang, Z.; Ji, Y. Comparative and phylogenetic analyses of the complete chloroplast genomes of three Arcto-Tertiary relicts: *Camptotheca acuminata*, *Davidia involucreata*, and *Nyssa sinensis*. *Front. Plant Sci.* **2017**, *8*, 1536. [[CrossRef](#)] [[PubMed](#)]
42. Wang, W.; Liu, H.; He, Q.; Yang, W.L.; Chen, Z.; Wang, M.; Su, Y.; Ma, T. Characterization of the complete chloroplast genome of *Camptotheca acuminata*. *Conserv. Genet. Resour.* **2017**, *9*, 241–243. [[CrossRef](#)]
43. Puterova, J.; Razumova, O.; Martinek, T.; Alexandrov, O.; Divashuk, M.; Kubat, Z.; Hobza, R.; Karlov, G.; Kejnovsky, E. Satellite DNA and transposable elements in seabuckthorn (*Hippophae rhamnoides*), a dioecious plant with small Y and large X chromosomes. *Genome Biol. Evol.* **2017**, *9*, 197–212. [[CrossRef](#)] [[PubMed](#)]
44. Rocha, E.P.C. An appraisal of the potential for illegitimate recombination in bacterial genomes and its consequences: From duplications to genome reduction. *Genome Res.* **2003**, *13*, 1123–1132. [[CrossRef](#)] [[PubMed](#)]
45. Kegel, A.; Martinez, P.; Carter, S.D.; Aström, S.U. Genome wide distribution of illegitimate recombination events in *Kluyveromyces lactis*. *Nucleic Acids Res.* **2006**, *34*, 1633–1645. [[CrossRef](#)] [[PubMed](#)]

46. Dodsworth, S.; Leitch, A.R.; Leitch, I.J. Genome size diversity in angiosperms and its influence on gene space. *Curr. Opin. Genet. Dev.* **2015**, *35*, 73–78. [[CrossRef](#)] [[PubMed](#)]
47. Lisch, D. How important are transposons for plant evolution? *Nat. Rev. Genet.* **2013**, *14*, 49–61. [[CrossRef](#)] [[PubMed](#)]
48. Zhong, B.; Yonezawa, T.; Zhong, Y.; Hasegawa, M. Episodic evolution and adaptation of chloroplast genomes in ancestral grasses. *PLoS ONE* **2009**, *4*, e5297. [[CrossRef](#)] [[PubMed](#)]
49. Shaw, J.; Lickey, E.B.; Schilling, E.E.; Small, R.L. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *Am. J. Bot.* **2007**, *94*, 275–288. [[CrossRef](#)] [[PubMed](#)]
50. Zhang, Z.; Gerstein, M. Patterns of nucleotide substitution, insertion and deletion in the human genome inferred from pseudogenes. *Nucleic Acids Res.* **2003**, *31*, 5338–5348. [[CrossRef](#)] [[PubMed](#)]
51. Matsuoka, Y.; Yamazaki, Y.; Ogihara, Y.; Tsunewaki, K. Whole chloroplast genome comparison of rice, maize, and wheat: Implications for chloroplast gene diversification and phylogeny of cereals. *Mol. Biol. Evol.* **2002**, *19*, 2084–2091. [[CrossRef](#)] [[PubMed](#)]
52. Castle, J.C. SNPs occur in regions with less genomic sequence conservation. *PLoS ONE* **2011**, *6*, e20660. [[CrossRef](#)] [[PubMed](#)]
53. Allegre, M.; Argout, X.; Boccara, M.; Fouet, O.; Roguet, Y.; Bérard, A.; Thévenin, J.M.; Chauveau, A.; Rivallan, R.; Clement, D. Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L. *DNA Res.* **2011**, *19*, 23–35. [[CrossRef](#)] [[PubMed](#)]
54. Wang, W.; Zhang, X. Identification of the sex-biased gene expression and putative sex-associated genes in *Eucommia ulmoides* Oliver using comparative transcriptome analyses. *Molecules* **2017**, *22*, 2255. [[CrossRef](#)] [[PubMed](#)]
55. Zhang, Y.-J.; Ma, P.-F.; Li, D.-Z. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS ONE* **2011**, *6*, e20596. [[CrossRef](#)] [[PubMed](#)]
56. Magee, A.M.; Aspinall, S.; Rice, D.W.; Cusack, B.P.; Sémon, M.; Perry, A.S.; Stefanović, S.; Milbourne, D.; Barth, S.; Palmer, J.D. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* **2010**, *20*, 1700–1710. [[CrossRef](#)] [[PubMed](#)]
57. Dong, W.; Xu, C.; Cheng, T.; Zhou, S. Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. *PLoS ONE* **2013**, *8*, e77965. [[CrossRef](#)] [[PubMed](#)]
58. Millen, R.S.; Olmstead, R.G.; Adams, K.L.; Palmer, J.D.; Lao, N.T.; Heggie, L.; Kavanagh, T.A.; Hibberd, J.M.; Gray, J.C.; Morden, C.W. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **2001**, *13*, 645–658. [[CrossRef](#)] [[PubMed](#)]
59. Guisinger, M.M.; Kuehl, J.V.; Boore, J.L.; Jansen, R.K. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18424–18429. [[CrossRef](#)] [[PubMed](#)]
60. Dugas, D.V.; Hernandez, D.; Koenen, E.J.; Schwarz, E.; Straub, S.; Hughes, C.E.; Jansen, R.K.; Nageswara-Rao, M.; Staats, M.; Trujillo, J.T. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Sci. Rep.* **2015**, *5*, 16958. [[CrossRef](#)] [[PubMed](#)]
61. Bock, D.G.; Kane, N.C.; Ebert, D.P.; Rieseberg, L.H. Genome skimming reveals the origin of the Jerusalem Artichoke tuber crop species: Neither from Jerusalem nor an artichoke. *New Phytol.* **2014**, *201*, 1021–1030. [[CrossRef](#)] [[PubMed](#)]
62. Downie, S.R.; Jansen, R.K. A comparative analysis of whole plastid genomes from the Apiales: Expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Syst. Bot.* **2015**, *40*, 336–351. [[CrossRef](#)]
63. Shaw, J.; Lickey, E.B.; Beck, J.T.; Farmer, S.B.; Liu, W.; Miller, J.; Siripun, K.C.; Winder, C.T.; Schilling, E.E.; Small, R.L. The tortoise and the hare II: Relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* **2005**, *92*, 142–166. [[CrossRef](#)] [[PubMed](#)]
64. Huang, J.; Yang, X.; Zhang, C.; Yin, X.; Liu, S.; Li, X. Development of chloroplast microsatellite markers and analysis of chloroplast diversity in Chinese jujube (*Ziziphus jujuba* Mill.) and wild jujube (*Ziziphus acidojujuba* Mill.). *PLoS ONE* **2015**, *10*, e0134519. [[CrossRef](#)] [[PubMed](#)]

65. Ren, X.; Jiang, H.; Yan, Z.; Chen, Y.; Zhou, X.; Huang, L.; Lei, Y.; Huang, J.; Yan, L.; Qi, Y. Genetic diversity and population structure of the major peanut (*Arachis hypogaea* L.) cultivars grown in China by SSR markers. *PLoS ONE* **2014**, *9*, e88091. [CrossRef] [PubMed]
66. Chen, Z.D.; Yang, T.; Lin, L.; Lu, L.M.; Li, H.L.; Sun, M.; Liu, B.; Chen, M.; Niu, Y.T.; Ye, J.F. Tree of life for the genera of Chinese vascular plants. *J. Syst. Evol.* **2016**, *54*, 277–306. [CrossRef]
67. Byng, J.W.; Chase, M.W.; Christenhusz, M.J.; Fay, M.F.; Judd, W.S.; Mabberley, D.J.; Sennikov, A.N.; Soltis, D.E.; Soltis, P.S.; Stevens, P.F. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 105–121.
68. Stevens, P.F. Angiosperm Phylogeny Website. Version 12, July 2012 (and More or Less Continuously Updated Since). Available online: <http://www.mobot.org/MOBOT/research/APweb/> (accessed on 1 March 2018).
69. Jansen, R.K.; Cai, Z.; Raubeson, L.A.; Daniell, H.; Depamphilis, C.W.; Leebensmack, J.; Müller, K.F.; Guisingerbellian, M.; Haberle, R.C.; Hansen, A.K. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 19369–19374. [CrossRef] [PubMed]
70. Smith, S.A.; Beaulieu, J.M.; Donoghue, M.J. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5897–5902. [CrossRef]
71. Davis, C.C.; Xi, Z.; Mathews, S. Plastid phylogenomics and green plant phylogeny: Almost full circle but not quite there. *BMC Biol.* **2014**, *12*, 11. [CrossRef] [PubMed]
72. Zeng, L.; Zhang, Q.; Sun, R.; Kong, H.; Zhang, N.; Ma, H.; Zeng, L.; Zhang, Q.; Sun, R.; Kong, H. Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* **2014**, *5*, 4956. [CrossRef] [PubMed]
73. Doyle, J.J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **1987**, *19*, 11–15.
74. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
75. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
76. Wyman, S.K.; Jansen, R.K.; Boore, J.L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **2004**, *20*, 3252–3255. [CrossRef] [PubMed]
77. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33* (Suppl. 2), W686–W689. [CrossRef] [PubMed]
78. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
79. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef] [PubMed]
80. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
81. Wang, W.; Ma, L.; Becher, H.; Garcia, S.; Kovarikova, A.; Leitch, I.J.; Leitch, A.R.; Kovarik, A. Astonishing 35S rDNA diversity in the gymnosperm species *Cycas revoluta* Thunb. *Chromosoma* **2016**, *125*, 683–699. [CrossRef] [PubMed]
82. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
83. Zhao, H.; Li, Y.; Peng, Z.; Sun, H.; Yue, X.; Lou, Y.; Dong, L.; Wang, L.; Gao, Z. Developing genome-wide microsatellite markers of bamboo and their applications on molecular marker assisted taxonomy for accessions in the genus *Phyllostachys*. *Sci. Rep.* **2015**, *5*, 8018. [CrossRef] [PubMed]
84. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map (SAM) format and SAMtools. *Transpl. Proc.* **2009**, *19*, 1653–1654.
85. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 1–33. [CrossRef]

86. Swofford, D.L. *PAUP*: Phylogenetic Analysis Using Parsimony, version 4.0 b10*; Sinauer Associates: Sunderland, MA, USA, 2003.
87. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **1985**, 783–791. [[CrossRef](#)] [[PubMed](#)]
88. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
89. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)] [[PubMed](#)]
90. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; pp. 1–8.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).