Research Article

Characterization of the Complete Chloroplast Genome of Acer truncatum Bunge (Sapindales: Aceraceae): A New Woody Oil Tree Species Producing Nervonic Acid

Qiuyue Ma,1 Yanan Wang,2 Lu Zhu,1 Changwei Bi,3 Shuxian Li,2 Shushun Li,1 Jing Wen,1 Kunyuan Yan,1 and Qianzhong Li1

1Institute of Leisure Agriculture, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China
2The Southern Modern Forestry Collaborative Innovation Center, Nanjing Forestry University, Nanjing, China
3School of Biological Science and Medical Engineering, Southeast University, Nanjing 210037, China

Correspondence should be addressed to Qianzhong Li; qianzhongli@jaas.ac.cn

Received 2 June 2019; Accepted 19 August 2019; Published 24 November 2019

Academic Editor: Shoichiro Ono

Copyright © 2019 Qiuyue Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Acer truncatum, which is a new woody oil tree species, is an important ornamental and medicinal plant in China. To assess the genetic diversity and relationships of A. truncatum, we analyzed its complete chloroplast (cp) genome sequence. The A. truncatum cp genome comprises 156,492 bp, with the large single-copy, small single-copy, and inverted repeat (IR) regions consisting of 86,010, 18,050, and 26,216 bp, respectively. The A. truncatum cp genome contains 112 unique functional genes (i.e., 4 rRNA, 30 tRNA, and 78 protein-coding genes) as well as 78 simple sequence repeats, 9 forward repeats, 1 reverse repeat, 5 palindromic repeats, and 7 tandem repeats. We analyzed the expansion/contraction of the IR regions in the cp genomes of six Acer species. A comparison of these cp genomes indicated the noncoding regions were more diverse than the coding regions. A phylogenetic analysis revealed that A. truncatum is closely related to A. miaotaiense. Moreover, a novel ycf4-cemA indel marker was developed for distinguishing several Acer species (i.e., A. buergerianum, A. truncatum, A. henryi, A. negundo, A. ginnala, and A. tonkinense). The results of the current study provide valuable information for future evolutionary studies and the molecular barcoding of Acer species.

1. Introduction

Acer truncatum Bunge, which is a member of the order Sapindales and the family Aceraceae, is a new versatile oil-producing woody tree that is widely distributed in northern China, Korea, and Japan, where it is a native species, but it has also been detected in Europe and North America [1]. This tree species represents a potential source of medicinal compounds. Many highly bioactive compounds have been extracted from Acer species, such as flavonoids, tannins, alkaloids, and terpenoids [2]. Acer truncatum seeds are processed to extract the seed oil, which was listed as a new food resource by the Ministry of Health of the People’s Republic of China in 2011. Approximately 5–6% of the A. truncatum seed oil is nervonic acid (C24:1) [3]. Nervonic acid, which is a key component of brain nerve cells and tissues, promotes the repair and regeneration of nerve cells and damaged tissues, and has been detected in the seed oil of a number of plants. Thus, A. truncatum seed oil represents a novel plant resource with potential applications for treating human cerebral and neurological problems [4].

Chloroplasts (cps) have important functions related to some essential metabolic pathways, including photosynthesis and glycometabolism [5, 6]. In plants, the DNA-replication mechanism associated with the cp genome is independent of the nuclear DNA-replication mechanism. Moreover, the cp genome is more highly conserved than the nuclear genome. In 1986, the liverwort (Marchantia polymorpha) cp genome became the first such genome to be described [7]. The subsequent emergence of rapid and cost-effective genome-sequencing technologies has led to more cp genomes being sequenced, with the resulting data deposited in the GenBank database. These sequences indicate that the angiosperm cp genomes typically form a circular DNA molecule comprising 120–170 kb that encode 120–130 genes [8]. The circular cp genome...
structure consists of the following four segments: two inverted repeat (IR) regions separated by large single-copy (LSC) and small single-copy (SSC) regions [9, 10]. However, genome size variations [11], rearrangement events [12–14], and gene losses [15] have been detected in some plant species. There is also considerable diversity in the IR size, possibly because the expansion and contraction of the IR regions have been very common events during the evolution of plant species, including those belonging to Fabaceae [16] and Poaceae [17]. The complete cp genome has been used in investigations of phylogenetic relationships, molecular markers, and evolution [18].

Insertions/deletions (indels) and single-nucleotide polymorphisms (SNPs) within the cp genome have been used to rapidly distinguish species [19–21]. Additionally, cp markers have been developed to identify closely related species, including buckwheat and the species of *Solanum*, *Angelica*, and other genera [20–22]. For example, Park et al. [23] used two indel markers (*trnK-trnQ* and *ycf1-ndhF*) to differentiate three *Aconitum* species. Additionally, indels in the *trnL-F*, *trnG-trnS*, and *trnL* introns have been used to analyze the molecular evolution of the *Silene* species cp genome [23]. Thus, indel and SNP cp markers are important for identifying species and investigating molecular evolution.

Several *Aceraceae* species have recently had their cp genome sequences published, including *Acer morrisonense* [24], *Dipteronia sinensis* and *Dipteronia dyeriana* [8], and *Acer griseum* [25]. Chen et al. [26] were the first to report the complete *A. truncatum* cp genome; however, they only focused on the genome composition and phylogenetic relationships. Thus, the *A. truncatum* cp genome was not comprehensively characterized. Compared with the result of Chen et al. [26], in our study, we also found *A. truncatum* is closely related to *A. miaotaiense*. Moreover, we also analyzed the repetition, contraction, and expansion of the IR regions as well as the synonymous and nonsynonymous substitution rates. Highly divergent regions and potential indels were detected via a comparative analysis of six available cp genome sequences. Additionally, on the basis of the results of our comparative analysis of cp genomes, we developed the *ycf4-cemA* indel marker to distinguish six *Acer* species (i.e., *A. buergerianum*, *A. truncatum*, *A. henryi*, *A. negundo*, *A. ginnala*, and *A. tonkinense*). The data presented herein will enrich the genetic information available for the genus *Acer*, provide novel insights into *A. truncatum* evolution, and form an important theoretical basis for increasing the *A. truncatum* seed yield.

2. Materials and Methods

2.1. DNA Sequencing and Chloroplast Genome Assembly. We collected fresh leaves from *A. truncatum* plants, which were obtained from the *Acer* germplasm collection of the Jiangsu Academy of Agricultural Sciences, Nanjing, Jiangsu, China. The leaves were frozen in liquid nitrogen and stored at −80°C. Total DNA was extracted from the frozen leaves with the DNA Isolation Kit (Aidlab, China). We prepared 350-bp shotgun libraries, which were sequenced according to the double-terminal sequencing method of the Illumina HiSeq X Ten platform.

A total of 16.30 GB high-quality clean data (Q30 > 95.23%) were used for assembling the sequence as described by Ferrarini et al. [27]. The cp DNA reads were extracted with SMALT, using the *A. buergerianum* (GenBank accession NC_034744), *A. miaotaiense* (GenBank accession NC_030343), and *A. morrisonense* (GenBank accession KT970611) cp genomes as queries. The reads with 90% similarity were considered to be derived from the cp genome. The data were trimmed with Sickle (https://github.com/najoshi/sickle) (using q = 30 as the threshold for trimming and l = 50 as the threshold for keeping a read based on length) and assembled with the default parameters of AbySS [28]. Redundant contigs were removed with the CD-Hit program [29] (threshold of 100%) and the unique contigs were merged with the default parameters of Minimus2. The boundary regions of LSC/IRB, IRB/SSC, SSC/IRA, and IRA/LSC of the completed cp genomes were validated with PCR-based sequencing. Details regarding the primers are provided in Supplementary Table S1.

2.2. Annotation and Comparative Analysis. The *A. truncatum* cp genome was annotated with DOGMA (http://dogma.ccbb.utexas.edu/). The start and stop codons were coupled manually. All tRNA genes were identified with the default parameters of ClustalW 2.0 [39]. The maximum likelihood (ML) analyses of phylogenetic relationships were completed among the 23 analyzed genomes as queries. The reads with 90% similarity were considered to be derived from the cp genome. The data were trimmed with Sickle (https://github.com/najoshi/sickle) (using q = 30 as the threshold for trimming and l = 50 as the threshold for keeping a read based on length) and assembled with the default parameters of AbySS [28]. Redundant contigs were removed with the CD-Hit program [29] (threshold of 100%) and the unique contigs were merged with the default parameters of Minimus2. The boundary regions of LSC/IRB, IRB/SSC, SSC/IRA, and IRA/LSC of the completed cp genomes were validated with PCR-based sequencing. Details regarding the primers are provided in Supplementary Table S1.

2.3. Analysis of Repeat Structures and Simple Sequence Repeats. Four types of repeat structures (i.e., forward repeat, palindromic repeat, reverse repeat, and complementary repeat) were identified with REPutter [35]. Additionally, tandem repeats were detected with the default settings of the Tandem Repeats Finder program (version 4.07b) [36]. The simple sequence repeats (SSRs) were analyzed with the MISA program. The motif size for mono-, di-, tri-, tetra-, penta-, and hexanucleotide SSRs was set as 10, 5, 4, 3, 3, and 3, respectively [37].

2.4. Analysis of Synonymous and Nonsynonymous Substitution Rates. The *A. truncatum*, *Citrus platymamma*, *Dimocarpus longan*, and *Spondias mombin* cp genome sequences were compared to determine the synonymous (*K*<sub>s</sub>) and nonsynonymous (*K*<sub>a</sub>) substitution rates. The protein-coding exons were separately aligned with MEGA 6.0. The *K*<sub>s</sub> and *K*<sub>a</sub> substitution rates were estimated with DnaSP [38].

2.5. Phylogenomic Analyses. A total of 22 whole cp genome sequences of Sapindales species (Supplementary Table S5) were used for elucidating the evolutionary status of *A. truncatum*, with *Euonymus hamiltonianus* (order Celastrales) serving as the outgroup. The 64 single-copy orthologous genes common among the 23 analyzed genomes were aligned with the default parameters of ClustalW 2.0 [39]. The maximum likelihood (ML) analyses of phylogenetic relationships were completed with RAxML using the GTRGAMMA model [40].
2.6. Estimation of the Divergence Time. For the divergence time, we first removed ambiguously aligned sites in the 23 whole genomes data set using GBLOCKS v.0.91b [41] with the following parameters: minimum sequences per conserved position, 15; minimum sequences per flank position, 20; maximum number of contiguous nonconserved positions, 8; minimum block length, 10; allowed gap positions, none. Then, the divergence time was estimated with the MCMCTree program of PAML (version 4.9a) [42], with the following parameters: burnin 100000, sampfreq 200, and nsample 10000. Moreover, E. hamiltonianus was constrained to be the outgroup, and the root age was constrained by the divergence time of E. hamiltonianus from A. truncatum (98–117 million years ago) (http://www.timetree.org/).

2.7. Development and Validation of the ycf4-cemA Indel Marker. The indel regions were selected based on the results of a similarity search with mVISTA. Additionally, primers were designed with Primer 5. The PCR amplification was performed as described by Ma et al. [43]. To confirm the accuracy of the PCR product sizes, three samples per species were sequenced by the General Biology Company (Nanjing, Jiangsu, China).

3. Results and Discussion

3.1. Features of the A. truncatum Chloroplast Genome. The A. truncatum genome sequence was submitted to the GenBank database (accession number MH638284). Chen et al. [26] was the first to describe the A. truncatum cp genomic features. Specifically, they reported that the A. truncatum cp genome comprises 156,262 bp, with an overall GC content of 37.9%. In the current study, we revealed similar structural features, with the A. truncatum cp genome consisting of 156,492 bp and forming a typical quadripartite structure (Figure 1 and Table 1). The LSC region (86,010 bp) and SSC region (18,050 bp) were separated by a pair of inverted repeats (IRA and IRB; 26,216 bp each). The GC content may be an important factor for assessing species similarity. The GC content of the complete A. truncatum cp genome was 37.90%, which was the same as the result of Chen et al. [26] and that of the LSC, SSC, and IR regions was 36.10%, 32.20%, and 42.80%, respectively, which is similar to the GC contents reported for other Acer species (Table 1) [24, 25]. The rRNA and tRNA genes had the highest GC contents in the IR regions across the complete cp genome, which is a phenomenon that is very common among plant species [44, 45].

We detected 134 genes in the A. truncatum cp genome, including 20 duplicated genes in the IR regions, 112 unique functional genes, and 2 pseudogenes. The 112 functional genes comprised 4 rRNA genes, 30 tRNA genes, and 78 protein-coding genes (Table 2). Among the 134 genes in the cp genome, 17 genes contained introns, of which three genes (ycf3, clpF, and rps12) contained two introns and the remaining genes contained one intron (i.e., eight protein-coding genes and six tRNA genes) (Table 2). The rps12 gene was trans-spliced, with its 3′ exon duplicated in the IRs and its 5′ exon located in the LSC region. Interestingly, trnK-UUU had the largest intron (2,487 bp) because of the presence of the matK gene. The infA and ycf1 genes were designated as pseudogenes. The infA gene contained several internal stop codons and the ycf1 gene was located at the boundary region of IR and SSC (Figure 1).

In this study, we assessed the relative synonymous codon usage (RSCU), which represents the nonuniform synonymous codon usage in coding sequences. Generally, RSCU values >1.00 and <1.00 indicate the codon is used more and less frequently than expected, respectively [46]. The codon usage frequency in the A. truncatum cp genome was estimated based on the protein-coding gene sequences (Table 3). The protein-coding genes comprised 77,796 bp encoding 25,932 codons. Leucine and cysteine were the most and least prevalent amino acids encoded by the codons, accounting for 10.82% and 1.17% of the codons, respectively. With the exception of the methionine and tryptophan codons, most of the amino acid codons had sequence biases [e.g., UUA (RSCU = 1.80) for leucine, UCU (RSCU = 1.56) for serine, and UAU (RSCU = 1.60) for tyrosine] (Table 3). Codon usage was generally biased toward A or T (U) with high RSCU values, which is a phenomenon that is very common among the cp genomes of land plant species [47, 48].

3.2. Analysis of the Repeats in the A. truncatum Chloroplast Genome. An analysis of the repeats in the A. truncatum cp genome revealed 22 long repeats (i.e., one reverse, nine forward, five palindromic, and seven tandem repeats). The only reverse repeat was 35 bp long. The forward and palindromic repeats were mainly longer than 30 bp (Supplementary Table S2 and Figure 2), whereas the tandem repeats were mainly 13–28 bp long (Supplementary Table S3). Most repeats were located in the intergenic spacers, with the rest located in protein-coding regions and introns. Short dispersed repeats are important for promoting cp genome rearrangements [49].

Simple sequence repeats are useful molecular markers for studying genetic diversity and identifying species [43]. In the

| Table 1: Details regarding the A. truncatum chloroplast genome. |
|---------------------------------------------------------------|
| **Characteristics** | **Chloroplast features** |
|--------------------|------------------------|
| Total cp DNA size (bp) | 156,492 |
| LSC size (bp) | 86,010 |
| SSC size (bp) | 18,050 |
| IR size (bp) | 26,216 |
| Total GC content (%) | 37.9 |
| GC content of LSC (%) | 36.1 |
| GC content of SSC (%) | 32.2 |
| GC content of IR (%) | 42.8 |
| Total CDS length (bp) | 77,796 |
| Protein-coding genes | 78 |
| tRNAs | 30 |
| rRNAs | 4 |
| Genes duplicated | 20 |
| Genes with a single intron(s) | 15 |
| Gene with two introns | 3 |
| Pseudogenes | 2 |
were AT. An examination of the distribution of the SSRs in the *A. truncatum* cp genome indicated that 73.08%, 21.79%, 3.85%, and 1.28% of the SSRs were in the intergenic spacer, protein-coding, intron, and tRNA regions, respectively (Figure 3(b)). Moreover, our data suggest that the *A. truncatum* cp genome contains fewer SSRs than the *A. miaotaiense* cp genome [24]. However, in both of these *Acer* species, the SSRs

**Figure 1:** Map of the complete *A. truncatum* chloroplast genome. Genes marked inside the circle are transcribed in a clockwise direction, and those on the outside of the circle are transcribed in a counter-clockwise direction. The dark and light gray areas correspond to the GC and AT contents, respectively.

In the current study, we detected 78 perfect microsatellites in the *A. truncatum* cp genome, including 67, 6, 1, and 4 mono-, di-, tri-, and tetranucleotide repeats, respectively; no hexanucleotide repeats were identified (Figure 3(a) and Supplementary Table S4). Most of these repeats were located in noncoding regions. Additionally, A or T accounted for 94.03% of the mononucleotide repeats, whereas all of the dinucleotide repeats were AT. An examination of the distribution of the SSRs in the *A. truncatum* cp genome indicated that 73.08%, 21.79%, 3.85%, and 1.28% of the SSRs were in the intergenic spacer, protein-coding, intron, and tRNA regions, respectively (Figure 3(b)). Moreover, our data suggest that the *A. truncatum* cp genome contains fewer SSRs than the *A. miaotaiense* cp genome [24]. However, in both of these *Acer* species, the SSRs
generally comprise A or T, which contributes to the A/T richness of their cp genomes. These results represent useful information regarding the cp SSR markers that can be applied to investigate the genetic diversity of *A. truncatum* as well as the relationships among species. These markers may also be relevant for selecting germplasms with high nervonic acid contents.

### 3.3. Contraction and Expansion of the IR Regions
The number and order of genes were highly conserved among the cp genomes of six Acer species. However, structural changes were detected in the IR boundaries (Figure 4). These changes represent a common evolutionary event and a major factor influencing the size differences among the cp genomes, implying they have an important evolutionary role in plants [50, 51]. We also compared the boundary regions of IR/LSC and IR/SSC in the cp genomes of *A. buergerianum, A. davidii, A. griseum, A. miaotaiense, A. morrisonense,* and *A. truncatum*. In the *A. buergerianum, A. miaotaiense,* and *A. truncatum* cp genomes, the ndhF genes were located in the LSC/IRb and SSC/IRb regions, respectively. The rpl22 genes were detected at the junctions of the LSC/IRb, SSC/IR, and LSC/IRa boundary regions, respectively (Figure 4). However, the rps19 gene was located entirely in the LSC region in the *A. miaotaiense* cp genome, but not in the other cp genomes. Additionally, in the *A. buergerianum* and *A. truncatum* cp genomes, the ycf1 gene was located in the SSC/IRa border regions, which resulted in a pseudogene in the IRb region. The cp genomes of the other three species (*A. davidii, A. griseum,* and *A. morrisonense*) exhibited a similar trend regarding the IR contraction and expansion. However, the rps19 gene was located in the SSC/IRb regions, respectively. The rpl22 gene extended 376bp into the IRb region. In all cp genomes, the trnH gene was located in the LSC region. Overall, we detected the contraction and expansion of the IR regions in all six analyzed Acer cp genomes.

### 3.4. Comparative Analysis of Six Acer Chloroplast Genomes
A comparative analysis of cp genomes is important for elucidating phylogenetic relationships and identifying species [52, 53]. With the annotated *A. morrisonense* cp genome as the reference, the overall sequence identities among the six analyzed Acer cp genomes were determined and visualized with mVISTA (Figure 5). The comparative cp genome analysis proved that the noncoding regions were more diverse than the coding regions, which is consistent with the findings in other plant species [54]. The IR regions were more conserved than the LSC and SSC regions, and four rRNA genes were essentially identical in the six Acer species. The intergenic spacers were relatively diverse (e.g., *trnH-psbA, matK-rps16, petN-psbM, petA-psbI,* and *ycf4-cemA*). The most diverse coding regions were the *matK, rps2, rpoC2, rpoB, rps19,* and *ycf1* sequences. Similar results were observed in previous studies [55, 56]. The highly diverse regions identified in the current study may be relevant for developing markers or genetic barcodes useful for exploring the genetic differentiation among Aceraceae species.

### 3.5. Analysis of Synonymous and Nonsynonymous Substitution Rates
In a previous study, the nonsynonymous and synonymous substitution ratio (*K*/*K*) was used to evaluate the evolutionary forces on some genes [49]. In this study, the *K*/*K* ratio was determined for 78 protein-coding genes following the comparison of the *A. truncatum* cp genome with the cp genomes of *C. platymamma, D. longan,* and *S. mombin* (Figure 6). Nearly all of the *K*/*K* ratios were less than 1.0, implying most of the protein-coding genes were under purifying selection during evolution. However, the *K*/*K* ratio of seven genes (*atpF, matK, psbD, rps16, rps18, rpl36, ndhB,* and *ycf1*) was between 0.5 and 1.0. Moreover, the *K*/*K* ratio was greater than 1 for *psal clpP, rps4, rpl22* and *ycf2,* which indicated these genes were under positive selection.

---

**Table 2: Gene composition of the *A. truncatum* chloroplast genome.**

| Classification of genes | Gene name | Number |
|-------------------------|-----------|--------|
| 1 | Photosystem I | psAA, B, C, I, J | 5 |
| 2 | ATP synthase | atpA, B, E, F, H, I | 6 |
| 3 | Photosystem II | psBA, B, C, D, E, F, H, I, J, K, M, N, T, Z | 15 |
| 4 | Rubisco | rbcL | 1 |
| 5 | Cytochrome | petA, B, D, G, L, N | 6 |
| 6 | NADH oxidoreductase | ndhAa, Bb(×2), C, D, E, F, G, H, I, J, K | 12 |
| 7 | Ribosomal proteins (SSU) | rps2, 3, 4, 7(×2), 8, 11, 12(×2)ac, 14, 15, 16, 18, 19c | 14 |
| 8 | Ribosomal proteins (LSU) | rpl22ac(×2), 14, 16, 20, 22, 23(×2), 32, 33, 36 | 11 |
| 9 | RNA polymerase | rpoA, rpoB, rpoC1a, rpoC2 | 4 |
| 10 | Ribosomal RNAs | rnr4.5c(×2), 5c(×2), 16c(×2), 23c(×2) | 8 |
| 11 | Other proteins | accD, ccsA, matK, clpPb, cemA | 5 |
| 12 | Transfer RNAs | trnA-UGC(×2)ac, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnF-MAU, trnG-UCc, trnG-GCC, trnH-GUG(×2), trnI-GAU (×2)ac, trnL-CAAc(×2), trnL-UAAc, trnL-UAG, trnM-CAU, trnN-GUU(×2)c, trnP-UGG, trnQ-UUG, trnK-UUUc, trnR-UUC, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-GUG, trnV-GAC(×2)c, trnR-ACG(×2)c, trnV-UACc, trnW-CCA, trnY-GUA, trnL-CAU(×2)c | 38 |
| 13 | Hypothetical proteins | ycf1, 2c(×2), 3c, 4, 15c(×2), | 7 |
| 14 | Pseudogenes | infA, ycf1 | 2 |
| Total | | | 134 |

*a* A single intron gene. *b* Two introns gene. *c* Two gene copies in IRs regions. *d* Gene divided into two independent transcription units. *e* Pseudogene.
during evolution. High $K_e/K_s$ ratios have been reported for some genes, including $ndhC$, $rps16$, and $ycf2$ [49]. These results clearly indicate that cp genes in different plant species may be subjected to diverse selection pressures.

3.6. Phylogenetic Analysis. Chloroplast genome sequences are valuable genomic resources for elucidating evolutionary history and have been widely applied in phylogenetic studies [55–59]. In the current study, to determine the phylogenetic position of $A. truncatum$, 22 complete cp genome sequences of Sapindales species were obtained from the GenBank database (Supplementary Table S5). A set of 64 single-copy orthologous genes present in the 23 analyzed cp genomes was used to construct phylogenetic trees, with $E. hamiltonianus$ serving as the outgroup. All Aceraceae species, including Acer and Dipteronia species, were grouped in one clade, which was consistent with the results of earlier investigations [25, 60, 61].
node (107.2 Mya), which is the divergence time of E. hamiltonianus from A. truncatum (98–117 million years ago) (http://www.timetree.org/). Results of divergence dates for some of the observed clades as well as the upper and lower bounds of the 95% highest posterior density intervals are shown on Figure 8. According to the MCMCTREE time estimates, the estimated divergence date for Burseraceae and Anacardiaceae, Meliaceae, and Simaroubaceae were 75.9 (52.9–95.8) Mya, and 73.2 (53.9–91.9) Mya, respectively. These results are in agreement with recent study [62]. Additionally, the Spaindaceae
and Aceraceae began to split at 64.4 (42.6–87.4) Mya. The divergence time of Acer from Dipterocarya is 14.7 (9.0–24.6) Mya within Aceraceae species. Divergence of A. buergerianum from a common ancestor with the five other Aceraceae species was estimated at 13.7 (8.3–23.2) Mya. Moreover, a recent divergence event between A. truncatum and A. miaotaiense around 1.6 (0.7–3.6) Mya. These results of our study will provide insights into the evolutionary relationships among Acer species.

3.8. Development of the ycf4-cemA Indel Marker. Because indel regions are relatively easy to detect, they are often used to develop markers for identifying species [63]. In the current study, the sequence variability of the large indel regions, which was revealed by sequence alignments with mVISTA, was used to develop markers. A comparison with the A. truncatum cp genome sequence detected a 91-bp deletion in the ycf4-cemA region of the A. buergerianum cp genome. The following six Acer species were selected to characterize the ycf4-cemA sequence: A. tonkinense, A. ginnala, A. negundo, A. henryi, A. truncatum, and A. buergerianum. To develop indel markers, sequence-specific primers were designed to anneal to the conserved regions flanking ycf4 and cemA (Table 4). The predicted products were successfully amplified with the ycf4-cemA-F/R primers for all 24 tested samples (Figure 9(a)). The length of the amplified ycf4-cemA sequence was similar for A. tonkinense, A. ginnala, A. negundo, A. henryi, A. truncatum, and A. buergerianum. In contrast, the corresponding sequence in A. buergerianum was shorter because of the 91-bp deletion (Figures 9(a) and 9(b)). As presented in Figure 9(a), A. tonkinense, A. ginnala, A. negundo, A. henryi, A. truncatum, and A. buergerianum yielded amplicons of 1,324, 1,320, 1,316, 1,314, and 1,235 bp, respectively. Two polythymine repeats were identified in the sequenced fragments. Interestingly, A. truncatum had an 8-bp insertion that was lacking in the other species. Other deletions are listed in

Figure 5: Comparison of the chloroplast genomes of six Acer species with mVISTA. The 50% and 100% refer to the similarity among sequences. Gray arrows above the aligned sequences represent genes and their orientation.
been identified using this approach. Thus, indel markers may represent an important resource for identifying species. The ycf4-cemA indel marker developed in this study may be applicable for species classifications and the identification of Acer species.

Supplementary Table S6. The predicted sizes of the indels were consistent with the sizes of the fragments amplified from the 24 samples analyzed in this study. Indel markers have commonly been used to distinguish closely related species in previous studies [22, 23]. However, Acer species have not been identified using this approach. Thus, indel markers may represent an important resource for identifying species. The ycf4-cemA indel marker developed in this study may be applicable for species classifications and the identification of Acer species.
Figure 7: Maximum-likelihood tree based on 64 single-copy orthologous genes shared among 23 species. Numbers at the nodes are bootstrap support values. The position of A. truncatum is indicated in bold. Euonymus hamiltonianus served as the outgroup.

Figure 8: Divergence times of 22 species of Sapindales based on complete chloroplast sequences. Values at nodes indicate divergence dates in millions of years. Euonymus hamiltonianus served as the outgroup.
Figure 9: Schematic diagram of the development of the ycf4-cemA indel marker in six Acer species. (a) Results of the PCR amplification of the ycf4-cemA indel marker in the following Acer species: (1) A. tonkinense, (2) A. ginnala, (3) A. negundo, (4) A. henryi, (5) A. truncatum, and (6) A. buergerianum; M, 2,000-bp ladder. (b) Alignment of the ycf4-cemA marker sequence with MEGA 6.0. The arrow indicates the 91-bp deletion in A. buergerianum.

Data Availability

The Acer truncatum chloroplast genome sequence was deposited in the GenBank database (accession MH638284).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Experimental design, Q.-Y.M.; collection and identification of plant materials, Y.-N.W.; L.Z., S.-X.L., J.W., S.-X.L., Q.-Z.L, and K.-Y.Y.; genome analysis, Q.-Y.M., Y.-N.W., and C.-W.B.; manuscript draft preparation, Q.-Y.M.; manuscript review, Q.-Y.M., S.-X.L., and Q.-Z.L. All authors contributed to the experiments and approved the final manuscript.

Acknowledgments

This research was supported by the Natural Science Foundation of China (31700628), the Natural Science Foundation of Jiangsu Province (BK20170602), the Independent Innovation Fund Project of Agricultural Science and Technology in Jiangsu Province (CX[17]1004), and the Technology Innovation and Extension Project of Forestry Science in Jiangsu Province (LYKJ[2018]14). We thank Liwen Bianji, Edanz Editing China (http://www.liwenbianji.cn/ac) for editing the English text of a draft of this manuscript.

Supplementary Materials

Supplementary Table S1: PCR-based sequence validation of junctions between the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IRA and IRb) regions of the A. truncatum chloroplast genome. Supplementary Table S2: Long repeat sequences in the A. truncatum chloroplast genome. Supplementary Table S3: Details regarding the tandem repeats in the A. truncatum chloroplast genome. Supplementary Table S4: Distribution of SSRs in the A. truncatum chloroplast genome. Supplementary Table S5: Details regarding the chloroplast genome sequences used for the phylogenetic analysis. Supplementary Table S6: The sequences variability of sequenced fragments. (Supplementary Materials)

References

[1] X. Guo, R. Wang, R. Chang et al., “Effects of nitrogen addition on growth and photosynthetic characteristics of Acer truncatum seedlings,” Dendrobiology, vol. 72, pp. 151–161, 2014.
[2] W. Tang, J. Wang, J. Xu, L. Wang, J. Huang, and Y. Chen, “Advances of chemical composition of medicinal plants in Aceraceae,” Northern Horticulture, vol. 36, pp. 194–200, 2012.

[3] X.-Y. Yang, J.-S. Fan, S.-Y. Wang, and R.-C. Sun, “A new resource of nervonic acid from purpleblow maple (Acer truncatum) seed oil,” Forest Products Journal, vol. 56, no. 11/12, pp. 147–150, 2006.

[4] R.-N. Yang, L.-X. Zhang, P.-W. Li et al., “A review of chemical composition and nutritional properties of minor vegetable oils in china,” Trends in Food Science & Technology, vol. 74, pp. 26–32, 2018.

[5] H. Daniell, C.-S. Lin, M. Yu, and W.-J. Chang, “Chloroplast genomes: diversity, evolution, and applications in genetic engineering,” Genome Biology, vol. 17, no. 1, Article ID 134, 2016.

[6] I. J. Tetlow, S. Rawsthorne, C. Raines, and M. J. Emes, “Plastid metabolic pathways,” In Annual Plant Reviews, Plastids, vol. 13, Blackwell, Hoboken, NJ, USA, S. G. Moller, Ed., pp. 60–125, 2009.

[7] K. Ohyama, H. Fukuzawa, T. Kohchi et al., “Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha chloroplast DNA,” Nature, vol. 322, no. 6079, pp. 572–574, 1986.

[8] B. R. Green, “Chloroplast genomes of photosynthetic eukaryotes,” Plant Journal, vol. 66, no. 1, pp. 34–44, 2011.

[9] T. Zhou, C. Chen, Y. Wei et al., “Comparative transcriptome and chloroplast genome analyses of two related Dipterona species,” Frontiers in Plant Science, vol. 7, 2016.

[10] H.-Y. Liu, Y. Yu, Y.-Q. Deng, J. Li, Z.-X. Huang, and S.-D. Zhou, “The chloroplast genome of Lilium henrici: genome structure and comparative analysis,” Molecules, vol. 23, no. 6, Article ID 1512, p. 1276, 2018.

[11] X.-M. Zheng, J. Wang, L. Feng et al., “Inferring the evolutionary mechanism of the chloroplast genome size by comparing whole-chloroplast genome sequences in seed plants,” Scientific Reports, vol. 7, no. 1, Article ID 1555, 2017.

[12] Y. Ogihara, T. Terachi, and T. Sasakuma, “Intronucleolar recombination of chloroplast genome mediated by short direct-repeat sequences in wheat species,” Proceedings of the National Academy of Sciences of the United States of America, vol. 85, no. 22, pp. 8573–8577, 1988.

[13] W. Wang, S. Chen, and X. Zhang, “Whole-genome comparison reveals heterogeneous divergence and mutation hotspots in chloroplast genome of Eucommia ulmoides oliver,” International Journal of Molecular Sciences, vol. 19, no. 4, p. E1037, 2018.

[14] C. Gu, L. Tembrock, S. Zheng, and Z. Wu, “The complete chloroplast genome of Catha edulis: a comparative analysis of genome features with related species,” International Journal of Molecular Sciences, vol. 19, no. 2, p. 525, 2018.

[15] W.-C. Wang, S.-Y. Chen, and X.-Z. Zhang, “Chloroplast genome evolution in actinidiaeae: cpP loss, heterogeneous divergence and phylogenomic practice,” PLoS One, vol. 11, no. 9, p. e0162324, 2016.

[16] S. Wicke, G. M. Schneeweiss, K. F. Muller, and D. Quandt, “The evolution of the plastid chromosome in land plants: gene content, gene order, gene function,” Plant Molecular Biology, vol. 76, no. 3–5, pp. 273–297, 2011.

[17] P.-F. Ma, Y.-X. Zhang, C.-X. Zeng, Z.-H. Guo, and D.-Z. Li, “Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (Poaceae),” Systematic Biology, vol. 63, no. 6, pp. 933–950, 2014.

[18] M. Parks, R. Cronn, and A. Liston, “Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes,” BMC Biology, vol. 7, no. 1, p. 84, 2009.

[19] K. Kim, S. C. Lee, J. Lee et al., “Comprehensive survey of genetic diversity in chloroplast genomes and 45S rDNAs within Panax ginseng species,” PLoS One, vol. 10, no. 6, p. e0117159, 2015.

[20] K.-S. Cho, B.-K. Yun, Y.-H. Yoon et al., “Complete chloroplast genome sequence of tartary buckwheat (Fagopyrum tataricum) and comparative analysis with common buckwheat (F. esculentum),” PLoS One, vol. 10, no. 5, p. e0125332, 2015.

[21] K. S. Cho, K. S. Cheon, S. Y. Hong et al., “Complete chloroplast genome sequences of Solanum commersonii and its application to chloroplast genotype in somatic hybrids with Solanum tuberosum,” Plant Cell Reports, vol. 35, no. 10, pp. 2113–2123, 2016.

[22] I. Park, S. Yang, W. J. Kim et al., “Sequencing and comparative analysis of the chloroplast genome of Angelica polymorpha and the development of a novel Indel marker for species identification,” Molecules, vol. 24, no. 6, p. 1038, 2019.

[23] P. K. Ingvarsson, S. Ribstein, and D. R. Taylor, “Molecular evolution of insertions and deletions in the chloroplast genome of silene,” Molecular Biology and Evolution, vol. 20, no. 11, pp. 1737–1740, 2003.

[24] Z.-H. Li, Y.-S. Xie, T. Zhou, Y. Jia, Y. L. He, and J. Yang, “The complete chloroplast genome sequence of Acer morrisonese (Aceraceae),” Mitochondrial DNA Part A, vol. 28, no. 3, pp. 309–310, 2015.

[25] W. C. Wang, S.-Y. Chen, and X.-Z. Zhang, “The complete chloroplast genome of the endangered Chinese paperback maple, Acer griseum, (Sapindaceae),” Conservation Genetics Resources, vol. 9, no. 4, pp. 527–529, 2017.

[26] S. Chen, B. Liu, S. Zhang, and J. Huang, “The complete chloroplast genome of Acer truncatum bunge (Aceraceae),” Mitochondrial DNA Part B, vol. 4, pp. 607–608, 2019.

[27] M. Ferrarini, M. Moretto, J. A. Ward et al., “An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome,” BMC Genomics, vol. 14, no. 1, p. 670, 2013.

[28] J. T. Simpson, K. Wong, and S. D. Jackman, “ABySS: a parallel assembler for short read sequence data,” Genome Research, vol. 19, pp. 1117–1123, 2009.

[29] W. Li and A. Godzik, “Cdh-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences,” Bioinformatics, vol. 22, no. 13, pp. 1658–1659, 2006.

[30] C. Liu, L. Shi, Y. Zhu et al., “CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences,” BMC Genomics, vol. 13, no. 1, p. 715, 2012.

[31] M. Lohse, O. Drehsel, and R. Bock, “OrganelleGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes,” Current Genetics, vol. 52, no. 5–6, pp. 267–274, 2007.

[32] K. Tamura, G. Stecher, D. Peterson, A. Filipski, and S. Kumar, “MEGA6: molecular evolutionary genetics analysis version,” Molecular Biology and Evolution, vol. 30, no. 12, pp. 2725–2729, 2013.

[33] K. A. Frazer, L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak, “VISTA: computational tools for comparative genomics,” Nucleic Acids Research, vol. 32, pp. W273–W279, 2004.

[34] C. Mayor, M. Brudno, J. R. Schwartz et al., “VISTA: visualizing global DNA sequence alignments of arbitrary length,” Bioinformatics, vol. 16, pp. 1046–1047, 2000.
The chloroplast genomes of herbaceous bamboos (Bambusoideae: Olyreae) have been extensively studied. For example, L. A. Raub et al. conducted a comparative analysis of sequence evolution among 17 vascular plant DNA Research sequences from Korean ginseng (Panax schinseng), which helped measure the directional synonymous codon usage bias and its potential applications. P. M. Sha et al. used the codon adaptation Index-a measure of directional synonymous codon usage bias, in its a/f_ter removing divergent and ambiguously aligned blocks. F. Zhang, S. Chen, and S. Li, among others, used the CodonW tool to analyze codon usage bias. Y. Du, Y. Bi, X. Chen, F. Yang, J. Xue, and X. Zhang, reported on the complete chloroplast genome of Liliaceae, its structure and evolution, in Conservation Genetics Resources, vol. 8, no. 4, pp. 375–378, 2016.

Kunyong Lee et al. (2014) used Clustal X version 2.0 to perform comparative analysis and post-analysis of large phylogenies. M. A. Gitzendanner, P. S. Soltis, G. K. Wong, B. R. Ruhfel, and D. E. Soltis, in their study on plastid phylogenetic analysis of green plants: a billion years of evolutionary history, published in American Journal of Botany, vol. 105, no. 3, pp. 291–301, 2018. They also contributed to the development of chloroplast genome analysis of two pangolin species: Manis pentadactyla and Manis javanica, in GigaScience, vol. 5, no. 1, pp. 1–5, 2016.

Comparative optical genome analysis of two pangolin species: Manis pentadactyla and Manis javanica, GigaScience, vol. 5, no. 1, pp. 1–5, 2016. X. Chi, J. Wang, Q. Gao, F. Zhang, and S. Chen, “The complete chloroplast genomes of two Lancea species with comparative analysis,” Molecules, vol. 23, no. 3, p. 602, 2018.

S. Zeng, T. Zhou, K. Han, Y. Yang, J. Zhao, and Z.-L. Liu, “The complete chloroplast genome sequences of six Rehmannia species,” Genes, vol. 8, no. 3, p. 103, 2017. S. Y. Hong, K. S. Cheon, K. O. Yoo et al., “Complete chloroplast genome sequences and comparative analysis of Chrysophyllum quinoa and C. album,” Frontiers in Plant Science, vol. 8, Article ID 1686, 2017.

M. A. Gitzendanner, P. S. Soltis, G. K. Wong, B. R. Ruhfel, and D. E. Soltis, “Plastid phylogenetic analysis of green plants: a billion years of evolutionary history,” American Journal of Botany, vol. 105, no. 3, pp. 291–301, 2018. Y. Du, Y. Bi, X. Chen, F. Yang, J. Xue, and X. Zhang, “The complete chloroplast genome of Liliaceae, its structure and evolution,” Conservation Genetics Resources, vol. 8, no. 4, pp. 375–378, 2016.

J.-L. Cao, D. Jiang, Z.-Y. Zhao et al., “Development of chloroplast genomic resources in Chinese Yam, Dioscorea polystachya,” BioMed Research International, vol. 2018, pp. 1–11, 2018. T. Zhou, C. Chen, Y. Wei et al., “Comparative transcriptome and chloroplast genome analyses of two related Dipteronia species,” Frontiers in Plant Science, vol. 7, 2016.

J. Zhao, Y. Xu, L. Xi, J. Yang, H. Chen, and J. Zhang, “Characterization of the chloroplast genome sequence of Acer miaotaiense: comparative and phylogenetic analyses,” Molecules, vol. 23, no. 7, p. 1740, 2018.

A. N. Mueller-Riehl, A. Weeks, J. W. Clayton, and S. Buerki, “Molecular phylogenetics and molecular clock dating of sapindales based on plastid rbcL, atpB and trnl-trnf DNA sequences,” Taxon, vol. 65, no. 5, pp. 1019–1036, 2016.

Z. Suo, Z. Jia, Q. Li et al., “Distinguishing Haloxyll persicum and H. Ammodendron (Haloxyll Bunge, Amaranthaceae) using DNA marker,” AASRI Procedia, vol. 1, pp. 305–310, 2012.