The complete chloroplast genome of *Stauntonia chinensis* and compared analysis revealed adaptive evolution of subfamily Lardizabaloideae species in China

Feng Wen 1†, Xiaozhu Wu 1†, Tongjian Li 1, Mingliang Jia 1, Xinsheng Liu 1 and Liang Liao 1

Abstract

**Background:** *Stauntonia chinensis* DC. belongs to subfamily Lardizabaloideae, which is widely grown throughout southern China. It has been used as a traditional herbal medicinal plant, which could synthesize a number of triterpenoid saponins with anticancer and anti-inflammatory activities. However, the wild resources of this species and its relatives were threatened by over-exploitation before the genetic diversity and evolutionary analysis were uncovered. Thus, the complete chloroplast genome sequences of *Stauntonia chinensis* and comparative analysis of chloroplast genomes of Lardizabaloideae species are necessary and crucial to understand the plastome evolution of this subfamily.

**Results:** A series of analyses including genome structure, GC content, repeat structure, SSR component, nucleotide diversity and codon usage were performed by comparing chloroplast genomes of *Stauntonia chinensis* and its relatives. Although the chloroplast genomes of eight Lardizabaloideae plants were evolutionary conserved, the comparative analysis also showed several variation hotspots, which were considered as highly variable regions. Additionally, pairwise Ka/Ks analysis showed that most of the chloroplast genes of Lardizabaloideae species underwent purifying selection, whereas 25 chloroplast protein coding genes were identified with positive selection in this subfamily species by using branch-site model. Bayesian and ML phylogeny on CCG (complete chloroplast genome) and CDs (coding DNA sequences) produced a well-resolved phylogeny of Lardizabaloideae plastid lineages.

**Conclusions:** This study enhanced the understanding of the evolution of Lardizabaloideae and its relatives. All the obtained genetic resources will facilitate future studies in DNA barcode, species discrimination, the intraspecific and interspecific variability and the phylogenetic relationships of subfamily Lardizabaloideae.

**Keywords:** Herbal medicine, Plastome, Adaptation, Positive selection, Phylogeny analyses

Background

Herbal medicine has been used as complementary and alternative treatments to augment existing therapies all over the world. The bioactive natural compounds extracted in herbal medicine may have the potential to form new drugs to treat a disease or other health conditions [1]. However, the wild resources of these plant species were on the verge of exhaustion by plundering exploitation with the increasing demand for herbal medicine with significant economic value [2]. Previous studies of herbal medicine species mainly concentrated on the cultivation and phytochemical studies. Whereas, few studies have described the genetic diversity and phylogenetic analysis. The germplasm, genetic and
genomic resources need to be developed as potential tools to better exploit and utilize these herbal medicine species [3]. In addition, a good knowledge of genomic information of these species could provide insights for conservation and restoration efforts. Therefore, the molecular techniques are required to analyze the genetic diversity and phylogenetic relationship of these plants.

Chloroplasts contain their own genome, composing of approximately 130 genes, which has a typical quadripartite structure consisting of one large single copy region (LSC), one small single copy region (SSC) and a pair of inverted repeats (IRs) in most plants [4–6]. Unlike nuclear genomes, the chloroplast genome is a highly conserved circular DNA with stable genome, gene content, gene order, and much lower substitution rates [7–10]. Recently, with the development of next generation sequencing, it has become relatively easy to obtain the complete chloroplast genome of non-model taxa [11–13]. Thus, complete chloroplast genome has been shown to be useful in inferring evolutionary relationships at different taxonomic levels as an accessible genetic resource [14, 15]. On the other hand, although the chloroplast genome is often regarded as highly conserved, some mutation events and accelerated rates of evolution have been widely identified in particular genes or intergenic regions at taxonomic levels [7, 16–18]. The complete chloroplast genome has been considered to be informative for phylogenetic reconstruction and testing lineage-specific adaptive evolution of plants.

Lardizabaloideae (Lardizabalaceae) comprising approximately 50 species in nine genera [19]. It’s a core component of Ranunculales and belongs to the basal eudicots. Most species of Lardizabaloideae were considered as herbal medicinal plants, which were widespread in China, except tribe Lardizabaleae (including genus Boquila and genus Lardizabala). Stauntonia chinensis DC., belonging to the subfamily Lardizabaloideae, is widely grown throughout southern China, including Jiangxi, Guangdong, and Guangxi provinces [20]. It has been frequently utilized in traditional Chinese medicine known as “Ye Mu Gua” due to its anti-nociceptive, anti-inflammatory, and anti-hyperglycemic characteristics [21–23]. In this study, we reported and characterized the complete chloroplast genome sequence of Stauntonia chinensis and compared it with another 38 chloroplast genomes of Ranunculales taxa previously published (including species from Berberidaceae, Circaeasteraceae, Eupteleaceae, Lardizabalaceae, Menispermaceae, Papaveraceae, and Ranunculaceae). Our results will be useful as a resource for marker development, species discrimination, and the inference of phylogenetic relationships for family Lardizabalaceae based on these comprehensive analyses of chloroplast genomes.

Results

The chloroplast genome of Stauntonia chinensis

We obtained 6.73 Gb of Illumina paired-end sequencing data from genomic DNA of Stauntonia chinensis. A total of 44,897,908 paired-end reads were retrieved with a sequence length of 150 bp, while a total of 41,809,601 of high-quality reads were used for mapping. The complete chloroplast DNA of Stauntonia chinensis. Was a circular molecule of 157,819 bp with typical quadripartite structure of angiosperms, which was composed of a pair of inverted repeats (IRA and IRB) of 26,143 bp each, separated by a large single copy (LSC) region of 86,545 bp and a small single copy (SSC) region of 18,988 bp (Fig. 1 and Table 1). The genome contained a total of 113 genes, including 79 unique protein-coding genes, 30 unique tRNA genes and 4 unique rRNA genes (Table 1). Of 113 genes, six protein-coding genes (rpl2, rpl23, ycf2, ndhB, rps7, and rps12), seven tRNA genes (trnL-CAA, trnL-CAA, trnV-GAC, trnA-UGC, trnR-ACG, trnN-GUU) and 4 rRNA genes (rrn16, rrn23, rrn4.5, rrn5) were duplicated in the IR regions. The Stauntonia chinensis chloroplast genome encoded a variety of proteins, which were mostly involved in photosynthesis and other metabolic processes, including large rubisco subunit, thylakoid proteins and subunits of cytochrome b/f complex (Table 2). Among the Stauntonia chinensis chloroplast genes, fifteen distinctive genes, including atpF, ndhA, ndhB, petB, petD, rpl2, rpl16, rpoC1, rps16, trnA-UGC, trnG-GCC, trnL-GAU, trnK-LUU, trnL-LIAA, and trnV-UAC harbored a single intron, and three genes (clpP, rps12 and ycf3) contained two introns (Table 3). The gene rps12 had trans-splicing, with the 5’-exon 1 located in the LSC region and the 3’-exons 2 and 3 and intron located in the IR regions. The overall G/C content was 38.67%, whereas the corresponding values of LSC, SSC, and IR regions were 37.1, 33.68, and 43.08%, respectively.

Codon usage bias pattern

It is generally acknowledged that codon usage frequencies varied among genomes, among genes, and within genes [24]. Codon preferences was often explained by a balance between mutational biases and natural selection for translational optimization [25–27]. Optimal codons help to increase both the efficiency and accuracy of translation [28]. The codon usage and relative synonymous codon usage (RSCU) values in the Stauntonia chinensis chloroplast genome was calculated based on protein-coding genes (Table 4). In total, 85 protein-coding genes in the Stauntonia chinensis chloroplast genome were encoded by 26,246 codons. Among the codons, the most frequent amino acid was leucine (2701 codons, 10.29%), while cysteine (310 codons, 1.18%) was the least abundant amino acid excluding the stop
codons. Similar to other angiosperm chloroplast genome, codon usage in the *Stauntonia chinensis* chloroplast genome was biased towards A and U at the third codon position, according to RSCU values (with a threshold of RSCU > 1) [29]. Further, the pattern of codon usage bias in the subfamily Lardizabaloideae and other species in Ranunculales were investigated (Fig. 2, Additional file 1). We found that two parameters (codon bias index, CBI and frequency of optimal codons, Fop) involved in codon usage bias were higher in Lardizabaloideae species than other species in Ranunculales.

**Repeats and microsatellites analyses**

Five type of repeat structures, including tandem, forward, palindromic, complement, and reverse repeats were identified using REPuter software in eight sequenced chloroplast genomes of Lardizabaloideae species. Overall, 23–40 repeat sequences were identified in each chloroplast genome, of which 3–9 tandem repeats, 7–17 forward repeats, and 11–17 palindromic repeats were separately detected, while few complement and reverse repeats were screened, for instance, only one complement repeat was predicted in *Holboellia angustifolia* (Fig. 3a). More than half of these repeats (72.5% at least)
## Table 1
Statistics of the chloroplast genomes of *Stauntonia chinensis* and seven other Lardizabaloideae species

| Species              | Access No. | Genome length (bp) | GC content (%) | LSC length (bp) | SSC length (bp) | IR length (bp) | Gene Number | Protein-coding tRNAs | rRNAs | No. of pseudogenes | GC3s content (%) |
|----------------------|------------|--------------------|----------------|-----------------|-----------------|----------------|-------------|----------------------|--------|---------------------|------------------|
| Akebia trifoliata    | KU204898   | 158,339            | 38.7           | 87,057          | 19,024          | 26,129         | 132         | 85                   | 37     | 8                   | 2.85             |
| Akebia quinata       | KX611091   | 157,817            | 38.7           | 86,543          | 18,988          | 26,143         | 132         | 85                   | 37     | 8                   | 2.85             |
| Archakebia apetala   | MK468518   | 157,929            | 38.7           | 86,630          | 19,001          | 26,149         | 132         | 85                   | 37     | 8                   | 2.85             |
| Decaisnea insignis   | KY200671   | 158,683            | 38.5           | 87,187          | 19,162          | 26,167         | 132         | 85                   | 37     | 8                   | 2.83             |
| Holboellia angustifolia | MN401677  | 157,797            | 38.7           | 86,543          | 18,972          | 26,141         | 132         | 85                   | 37     | 8                   | 2.85             |
| Holboellia latifolia | MH394378   | 157,818            | 38.7           | 86,567          | 18,971          | 26,140         | 132         | 85                   | 37     | 8                   | 2.85             |
| Sinofranchetia chinensis | MK533615  | 158,015            | 38.4           | 86,324          | 18,923          | 26,384         | 133         | 85                   | 38     | 8                   | 2.81             |
| Stauntonia chinensis | MN401678   | 157,819            | 38.7           | 86,545          | 18,988          | 26,143         | 132         | 85                   | 37     | 8                   | 2.85             |
had a repeat length between 30 and 50 bp (Fig. 3b), and majority of the repeats were distributed in non-coding regions, including the intergenic regions and introns. Nevertheless, a small number of coding genes and tRNA genes were also found to contain repeat sequences, such as ycf2, psaA, psaB, trnG and trnS in *Stauntonia chinensis* chloroplast genome.

A total of 47–83 microsatellites were predicted in these eight chloroplast genomes, and the most predominant type of the SSRs were mononucleotides SSRs (especially for A/T, Fig. 3c). Besides, di-nucleotides were also detected in each chloroplast genomes, especially for AT5 and AT6. Furthermore, *Stauntonia chinensis* chloroplast genome contained four tri-nucleotides and four tetra-nucleotides, while other seven chloroplast genomes were found to have 34 tri-nucleotides and 31 tetra-nucleotides. Additionally, none of penta- and hexa-nucleotides were found in *Stauntonia chinensis* chloroplast genome.

### Table 3 Genes with introns in the chloroplast genome of *Stauntonia chinensis*

| Gene       | Location | Exon I (bp) | Extron I (bp) | Exon II (bp) | Extron II (bp) | Exon III (bp) |
|------------|----------|-------------|---------------|--------------|---------------|--------------|
| trnK-UUU   | LSC      | 37          | 2476          | 35           |               |              |
| rps16      | LSC      | 40          | 874           | 227          |               |              |
| trnG-UCC   | LSC      | 23          | 712           | 49           |               |              |
| atpF       | LSC      | 145         | 1088          | 71           |               |              |
| rpoC1      | LSC      | 432         | 764           | 1614         |               |              |
| ycf3       | LSC      | 124         | 721           | 230          | 737           | 153          |
| trnL-UAA   | LSC      | 35          | 508           | 50           |               |              |
| trnV-UAC   | LSC      | 39          | 587           | 35           |               |              |
| ctpP       | LSC      | 71          | 798           | 291          | 653           | 247          |
| petB       | LSC      | 6           | 807           | 642          |               |              |
| petD       | LSC      | 8           | 709           | 475          |               |              |
| rpl16      | LSC      | 9           | 1102          | 399          |               |              |
| rpl2       | IR       | 391         | 664           | 434          |               |              |
| ndhB       | IR       | 777         | 696           | 756          |               |              |
| rps12a     | IR       | 114         | 232           | 659          | 23           |
| trnL-GAU   | IR       | 37          | 939           | 35           |               |              |
| trnA-UAG   | IR       | 38          | 800           | 35           |               |              |
| ndhA       | SSC      | 553         | 1086          | 539          |               |              |

*^rps12 gene is trans-spliced gene with the two duplicated 3’ end exons in IR regions and 5’ end exon in the LSC region*
Similarly, SSRs mainly located in non-coding regions, particularly in intergenic regions, while several coding genes and tRNA genes such as trnK, trnG, ycf3, trnL, ndhK, cema, and ycf1 were also found to contain SSRs, especially, ycf1 has three types of SSRs.

**Genome comparison**

The border regions and adjacent genes of chloroplast genomes were compared to analyze the expansion and contraction variation in junction regions, which were common phenomena in the evolutionary history of land plants. To evaluate the potential impact of the junction changes, we compared the IR boundaries of the Lardizabaloideae species (Fig. 4). Although the majority of genomic structure, such as gene order and gene number were conserved, the eight chloroplast genomes of Lardizabaloideae species showed visible divergences at the IRA/LSC and IRB/SSC borders. Some differences in the IR expansions and contractions still existed. For example, the IRB region expanded into the gene rps19 with 87 and 250 bp in the IRB regions of Decaisnea insignis and Sinofranchetia chinensis chloroplast genomes, respectively, although the IRB regions of other six chloroplast genomes were conserved. Thus, we found that the IR regions of the eight chloroplast genomes were conserved, except the chloroplast genomes of Decaisnea.

### Table 4 Relative synonymous codon usage (RSCU) in the Stauntonia chinensis chloroplast genome

| Codon | Amino acid | Count | RSCU | Codon | Amino acid | Count | RSCU |
|-------|------------|-------|------|-------|------------|-------|------|
| UUU   | F          | 849   | 1.18 | UAU   | Y          | 766   | 1.61 |
| UUC   | F          | 587   | 0.82 | UAC   | Y          | 185   | 0.39 |
| UUA   | L          | 743   | 1.65 | UAA   | *          | 38    | 1.34 |
| UUG   | L          | 570   | 1.27 | UAG   | *          | 25    | 0.88 |
| CUU   | L          | 586   | 1.3   | CAU   | H          | 507   | 1.51 |
| CUC   | L          | 206   | 0.46 | CAC   | H          | 164   | 0.49 |
| CUA   | L          | 383   | 0.85 | CAA   | Q          | 688   | 1.48 |
| CUG   | L          | 213   | 0.47 | CAG   | Q          | 239   | 0.52 |
| AUU   | I          | 1055  | 1.42 | AAU   | N          | 945   | 1.52 |
| AUC   | I          | 487   | 0.66 | AAC   | N          | 297   | 0.48 |
| AUA   | I          | 682   | 0.92 | AAA   | K          | 956   | 1.43 |
| AUG   | M          | 627   | 1     | AAG   | K          | 379   | 0.57 |
| GUU   | V          | 521   | 1.44 | GAU   | D          | 865   | 1.57 |
| GUC   | V          | 179   | 0.49 | GAC   | D          | 237   | 0.43 |
| GUA   | V          | 521   | 1.44 | GAA   | E          | 977   | 1.45 |
| GUG   | V          | 227   | 0.63 | GAG   | E          | 366   | 0.55 |
| UCU   | S          | 543   | 1.57 | UGU   | C          | 219   | 1.41 |
| UCC   | S          | 361   | 1.04 | UGC   | C          | 91    | 0.59 |
| UCA   | S          | 443   | 1.28 | UGA   | *          | 22    | 0.78 |
| UCG   | S          | 211   | 0.61 | UGG   | W          | 455   | 1    |
| CUC   | P          | 419   | 1.52 | CGU   | R          | 360   | 1.33 |
| CCC   | P          | 220   | 0.8  | CGC   | R          | 100   | 0.37 |
| CCA   | P          | 330   | 1.2  | CGA   | R          | 363   | 1.34 |
| CCG   | P          | 134   | 0.49 | CGG   | R          | 118   | 0.43 |
| ACU   | T          | 521   | 1.53 | AGU   | S          | 390   | 1.13 |
| ACC   | T          | 261   | 0.76 | AGC   | S          | 125   | 0.36 |
| ACA   | T          | 422   | 1.24 | AGA   | R          | 499   | 1.84 |
| ACG   | T          | 161   | 0.47 | AGG   | R          | 188   | 0.69 |
| GCU   | A          | 630   | 1.8  | GGU   | G          | 608   | 1.34 |
| GCC   | A          | 213   | 0.61 | GGC   | G          | 179   | 0.39 |
| GCA   | A          | 400   | 1.14 | GGA   | G          | 734   | 1.62 |
| GCG   | A          | 160   | 0.46 | GGG   | G          | 296   | 0.65 |
Fig. 2 Statistics of codon usage bias in Lardizabaloideae and other family species. **a** CAI (Codon adaptation index), **b** CBI (Codon bias index), **c** FOP (Frequency of optimal codons index), **d** NC (Effective number of codons), **e** GC (GC content), **f** GC3s (GC of synonymous codons in 3rd position)
Fig. 3 The repeat elements in the chloroplast genome of eight Lardizabaloideae species. a Number of five repeat types; b Number of repeat sequences by length; c Number of six SSR types; d Number of identified SSR motifs.

Fig. 4 The comparison of the LSC, IR, and SSC boundary regions among the eight Lardizabaloideae species chloroplast genomes. φ indicates a pseudogene.
insignis and Sinofranchetia chinensis, which were slightly expanded compared with that of the other species.

To further investigate the divergence of chloroplast genomes among Lardizabaloideae species, a global sequence alignment of eight chloroplast genomes were compared using the annotated chloroplast genome of A. trifoliata as a reference. These closely related species had little difference in genome size, ranging from 157,797 bp to 158,683 bp. Although sequence similarities were very high in IR regions, the chloroplast genomes exhibited less conserved in LSC and SSC regions (Fig. 5).

A sliding window analyses of the whole chloroplast genomes of eight Lardizabaloideae species indicated that most of the variation occurred in the LSC and SSC regions, which exhibited higher nucleotide variability (Pi) in comparison to IR regions (Fig. 6a). As shown in Fig. 6a, the nucleotide diversity values in the LSC and SSC regions ranged from 0.00173 to 0.08625 and from 0.0044 to 0.05637, respectively, while the value was from 0.00 to 0.01131 in the IRs regions. Expectedly, the divergence in intergenic regions was higher than in genic regions, but the ycf1 gene exhibited a higher variability. The most divergent non-coding regions among the eight Lardizabaloideae chloroplast genomes were trnH-psbA, trnK-rps16, rps16-trnQ, trnC-petN, trnT-psbD, ycf3-trnS-rps4, trnT-trnL, accD-psaI, petA-psbl, ndhF-rpl32, and rpl32-trnL. Although coding regions were conserved, minor sequence variation was observed among the eight chloroplast genomes in the trnK, matK, psaJ, rpl16, ndhF, ccsA, ndhA, and ycf1 gene as shown in Fig. 6b (Pi value > 0.015). Similarly, mauve alignment results revealed that no large structural changes such as gene order rearrangements was detected across these eight chloroplast genomes of Lardizabaloideae species (Additional file 2), although some inversions were present in LSC and SSC regions in other Ranunculales species, such as Pulsatilla chinensis, Anemone trullifolia, and Anemoclema glaucifolium.

**Estimating rates of chloroplast evolution and positive selection analyses**

Most of Ka/Ks values of these Ranunculales species were less than or close to 1, providing the evidence that these chloroplast genes experienced purifying or no selection pressures (Fig. 7 and Additional file 3). Furthermore, in Lardizabaloideae species, the Ka/Ks ratios were far less than 1 among Akebia trifoliata, Akebia quinata, Stauntonia chinensis, and Archakebia apetala. However, the Ka/Ks ratio between Holboellia angustifolia and Holboellia latifolia was greater than 1, implying some
chloroplast coding sites of these two species were under positive selection.

To further identify chloroplast protein-coding genes that might have undergone positive selection in Lardizabaloideae species, branch-site model analysis was employed by defining Lardizabaloideae species as foreground branch. A total number of 55 single-copy coding genes were considered for the positive selection analysis (Table 5). Although the likelihood ratio test showed that most of p-values were not significant in each gene range, two protein coding genes (\textit{rbcL} and \textit{accD}) indicated rejection of a null model ($p < 0.05$), corroborating the hypothesis that some amino acid sites in these two proteins in clade Lardizabaloideae species have been under positive selection (Table 5). Further analysis using a Bayes empirical Bayes (BEB) procedure identified 25 protein coding genes (\textit{accD}, \textit{atpA}, \textit{atpE}, \textit{atpI}, \textit{ccsA}, \textit{clpP}, \textit{ndhD}, \textit{ndhF}, \textit{ndhH}, \textit{ndhJ}, \textit{psaA}, \textit{psaB}, \textit{psaI}, \textit{psbA}, \textit{psbZ}, \textit{rbcL}, \textit{rpl33}, \textit{rpoA}, \textit{rpoB}, \textit{rpoC1}, \textit{rps14}, \textit{rps2}, and \textit{ycf3}) with significant posterior probabilities suggesting some sites in these genes were under positive selection (Table 5, Fig. 8 and Additional file 4). Among them, 11 genes only had one positively selected site, whereas \textit{accD} gene contained the largest number of positively selected sites (16 sites). Notably, most of \textit{ndh} family genes possessed at least one positively selected site, implying this family members were potentially under positive selective pressure in Lardizabaloideae species (Fig. 8).

**Phylogenetic analyses**

Bayesian and ML trees reconstructed based on the CCG dataset were highly congruent in identifying the phylogenetic position of these seven families in the order Ranunculales (Fig. 9). All nodes of these phylogenetic trees were strongly supported by bootstrap values (BS) in ML analysis and posterior probabilities (PP) in Bayesian analysis. The 39 taxa were classified into five major clades, of which Berberidaceae, Menispermaceae, and Ranunculaceae species clustered into a clade showed a close genetic relationship, while other family species constituted a monophyly. However, the family Circaestereaceae species showed different position relative to other six families in Bayesian and ML reconstructed trees based on the protein-coding genes CDs dataset. The family Circaestereaceae species were clustered into a clade with family Ranunculaceae species in phylogenetic tree based on CDs dataset, indicating that Circaestereaceae had strong support to be a sister to the Ranunculaceae.
Discussion

Architecture of chloroplast genomes in subfamily Lardizabaloideae

Recently, chloroplast genomes have become to be useful tools to evaluate the genetic divergence among related species [30, 31]. Here we present the complete chloroplast genome of *Stauntonia chinensis*. The organization of the chloroplast genomes among eight Lardizabaloideae species exhibited a high degree of synteny, implying that these genomes were evolutionarily conserved at the genome-scale level (Table 1, Figs. 3, 5 and Additional file 5). On the contrary, there were still a few diverged coding genes, including *matK, accD, psaA, rpl16, ndhF, ycf1*, and so on. The *matK* and *ycf1* coding regions had been observed to be highly divergent and could serve as markers for DNA barcoding and phylogenetic analysis [32–35]. Similarly, nucleotide diversity analysis showed that eight genes (*trnK, matK, psaA, rpl16, ndhF, ccsA, ndhA*, and *ycf1*) among eight Lardizabaloideae species had higher divergence values (Pi > 0.015), implying that they contained more variations than other coding genes (Fig. 6b). Among these genes, *matK, ndhF, ccsA*, and *ycf1* have been
| Gene name | Null hypothesis | Alternative hypothesis | Significance test |
|-----------|-----------------|------------------------|-------------------|
|           | lnL  df Omega (w = 1) | lnL  df Omega (w > 1) | BEB               |
| accD      | −6244.2703 76 1 | −6251.82 75 40.32412 | 78 | 0.946,79 D 0.840,80 S 0.573,81 G 0.779,82 K 0.969*,87 G 0.750,90 D 0.597,148 M 0.557,162 P 0.654,167 I 0.715,188 N 0.626,208 S 0.753,262 N 0.683,295 N 0.585,314 V 0.694 |
| atpA      | −7235.2876 80 1 | −7235.29 79 1 | 123 V 0.784 |
| atpB      | −6617.3383 80 1 | −6617.34 79 1.00000 | 1.00E+00 |
| atpE      | −2016.8158 80 1 | −2018 79 87.46955 | 35 P 0.917,52 Y 0.977* |
| atpF      | −857.07265 80 1 | −857.073 79 1.00000 | 1.00E+00 |
| atpH      | −800.09979 76 1 | −800.1 75 3.48674 | 1.23E-01 |
| atpl      | −3183.0473 80 1 | −3183.51 79 20.22025 | 35 V 0.753,59 R 0.824,202 V 0.828 |
| ccsA      | −5988.5949 80 1 | −5989.75 79 15.32850 | 1.28E-01 |
| cemA      | −4097.1783 79 1 | −4097.63 78 999.00000 | 3.41E-01 |
| clpP      | −3779.5729 78 1 | −3779.64 77 1.73150 | 7.05E-01 |
| matK      | −9319.3607 80 1 | −9319.36 79 1.00000 | 1.00E+00 |
| ndhA      | −5632.3781 78 1 | −5632.38 77 1.00000 | 1.00E+00 |
| ndhC      | −1481.7425 79 1 | −1481.74 78 1.00000 | 9.99E-01 |
| ndhD      | −6748.038 78 1 | −6748.06 77 1.50673 | 8.29E-01 |
| ndhE      | −1550.4989 80 1 | −1550.5 79 1.00000 | 1.00E+00 |
| ndhF      | −14,665.344 76 1 | −14,665.3 75 1.00000 | 1.00E+00 |
| ndhG      | −2862.0535 79 1 | −2862.05 78 1.00000 | 1.00E+00 |
| ndhH      | −5898.3248 78 1 | −5898.32 77 1.00000 | 1.00E+00 |
| ndhl      | −2246.7819 78 1 | −2247.13 77 13.49799 | 4.02E-01 |
| ndhJ      | −2064.866 80 1 | −2064.87 79 1.00000 | 1.00E+00 |
| ndhK      | −2624.0218 76 1 | −2624.02 75 1.00000 | 1.00E+00 |
| petA      | −2607.1562 80 1 | −2607.16 79 10.70011 | 9.89E-01 |
| petG      | −404.32829 80 1 | −404.36 79 1.00000 | 7.76E-01 |
| petL      | −386.63805 80 1 | −385.507 79 1.41918 | 1.33E-01 |
| petN      | −218.82071 80 1 | −218.82 79 3.96657 | 9.95E-01 |
| psaA      | −7414.7714 80 1 | −7414.77 79 1.15065 | 9.85E-01 |
| psaB      | −7487.909 80 1 | −7487.91 79 1.00000 | 9.86E-01 |
| psaC      | −965.2544 80 1 | −965.254 79 1.00000 | 9.99E-01 |
| psal      | −489.20384 80 1 | −489.204 79 1.00000 | 9.94E-01 |
| psaJ      | −623.17909 80 1 | −622.516 79 1.00000 | 1.00E+00 |
| psbA      | −3805.4481 80 1 | −3805.45 79 1.00000 | 1.00E+00 |
| psbB      | −5464.4744 80 1 | −5464.47 79 1.00000 | 1.00E+00 |
| psbD      | −3430.3571 80 1 | −3430.36 79 1.00000 | 9.89E-01 |
| psbE      | −824.21551 80 1 | −824.216 79 1.00000 | 1.00E+00 |
| psbF      | −297.73519 80 1 | −297.735 79 1.00000 | 1.00E+00 |
| psbl      | −386.8387 80 1 | −386.839 79 1.00000 | 1.00E+00 |
| psbJ      | −409.58353 80 1 | −409.584 79 3.05096 | 9.99E-01 |
| psbK      | −808.36136 80 1 | −808.361 79 1.00000 | 1.00E+00 |
| psbL      | −251.14656 80 1 | −250.996 79 1.00000 | 5.83E-01 |
| psbM      | −344.18423 80 1 | −344.184 79 1.00000 | 1.00E+00 |
| psbN      | −339.98529 78 1 | −339.985 77 3.88417 | 1.00E+00 |
previously detected as highly variable regions in different plants, and some of those were served as DNA barcode [36–39]. However, previous studies confirmed that both introns and intergenic regions exhibited higher divergence levels than coding regions [40]. In our study, both genome-scale level alignments and nucleotide diversity analyses of the eight Lardizabaloideae chloroplast genomes revealed common variable sites, including eleven intergenic regions and eight coding genes (Figs. 5 and 6).

Table 5 The potential positive selection test based on the branch-site model (Continued)

| Gene name | Null hypothesis | Alternative hypothesis | Significance test |
|-----------|----------------|------------------------|------------------|
|           | lnL df Omega (w = 1) | lnL df Omega (w > 1) | BEB P-value |
| psbZ      | -615.4539 76 1 | -615.454 75 1.00000 | 17 L 0.553 |
| rbcL      | -6385.4416 80 1 | -6388.64 79 999.00000 | 365 L 0.908 |
| rpl20     | -2104.3884 80 1 | -2104.39 79 1.00000 | 1.14E-02 |
| rpl33     | -895.54775 80 1 | -895.771 79 84.08315 | 57 G 0.873 |
| rpoA      | -5786.0327 76 1 | -5786.03 75 1.00000 | 5 A 0.523,107 Y 0.502,125 S 0.554,135 Q 0.559,276 K 0.555 |
| rpoB      | -14,722,031 80 1 | -14,722 79 1.24476 | 281 A 0.535,454 Y 0.537 |
| rpoC1     | -9096.3909 78 1 | -9096.39 77 1.00000 | 110 A 0.561,590 P 0.569 |
| rpoC2     | -21,869,149 80 1 | -21,869,1 79 1.00000 | 1.00E+00 |
| rps14     | -1369.7427 78 1 | -1370.22 77 999.00000 | 28 R 0.844,35 L 0.559 |
| rps15     | -3240.00056 74 1 | -323,269 73 2.04455 | 2.26E-01 |
| rps18     | -1134.1501 80 1 | -1134.15 79 1.00000 | 1.00E+00 |
| rps2      | -3349.317 80 1 | -3349.32 79 1.00000 | 124 E 0.631,182 I 0.648 |
| ycf3      | -1601.2056 78 1 | -1601.35 77 24.41198 | 5 R 0.733 |
| ycf4      | -2660.8473 79 1 | -2661.13 78 71.02868 | 4.55E-01 |

*Indicate that the posterior probabilities of the site is > 0.95

Fig. 8 Partial alignment of six out of 25 positively selected genes. a-f Partial aligned amino acids sequences of the ndhD, ndhF, ndhH, ndhI, ndhJ and ndhK gene, respectively. The red blocks stand for the amino acids in Lardizabaloideae with a high BEB posterior probability.
Previous studies supported that repetitive sequences were considered to play crucial roles in chloroplast genome arrangement and sequence divergence, even those were generally rare among angiosperm plastomes [41–43]. Generally, Lardizabaloideae species exhibited a significant difference in number and length of repeats within their chloroplast genomes. Most of the repeats were distributed in non-coding regions, including the intergenic regions and introns, reflecting the fact that the evolution of non-coding regions was higher than that of coding regions (Fig. 3) [44]. However, several repeats occurred in the same gene (ycf2) or paralogs (pasA/psaB and trnS-GCU/trnS-UGA/trnS-GGA), which might be caused by replication slippage, generating improper sequence recombination [45, 46]. Because of analytical and highly polymorphic nature, SSRs were considered to be well suited to assessment of genetic diversity within species and their relatives [47, 48]. In summary, repetitive sequences present in chloroplast genomes could facilitate the species discrimination and act as tools for investigating levels of genetic diversity in subfamily Lardizabaloideae.

**The adaptive evolution and positive selection**

The Ka/Ks ratios were important to deduce the evolutionary rates and understand the adaptive developments among species [49]. The pairwise Ka/Ks ratios among Akebia trifoliata, Akebia quinata, Stauntonia chinensis, and Archakebia apetala were far less than 1, suggesting more intense purifying selection in these species, for both conservative and radical nonsynonymous substitutions (Fig. 7, Additional file 3). The lower Ka/Ks ratios might be explained that most genes in these species were likely to undergo deleterious nonsynonymous substitutions, and the purifying selection with stronger selective constraints for nonsynonymous substitutions than for synonymous ones [50, 51]. However, the Ka/Ks ratio between H. angustifolia and H. latifolia was greater than 1, implying some chloroplast coding sites of these two species were under positive selection. It is possible that more unknown selective forces might have contributed to the elevated Ka/Ks ratios, and resulted in species divergence [52].

It was suggested that codon sites with higher posterior probability could be also considered as positively selected sites, and genes containing positively selected sites might be evolving under divergent selective pressures [53, 54]. Although pairwise Ka/Ks ratios showed most of the chloroplast genes of Ranunculales species experienced purifying or no selection pressures, at least 25 chloroplast protein coding genes were identified with significant posterior probabilities suggesting sites with positive selection in Lardizabaloideae species, which indicated these genes might have evolved to adapt to environmental conditions (Table 5). Notably, we found that five of these 25 genes were associated with
photosystem I and II subunits (*psaA, psaB, psaI, psbA*, and *psbZ*), while six of ten NADH-dehydrogenase subunit genes (*ndhD, ndhF, ndhH, ndhI, ndhJ*, and *ndhK*) possessed at least one positively selected site, implying these family members were potentially under positive selective pressure in Lardizabaloideae species (Fig. 8). Photosystem subunits and NADH-dehydrogenase subunits were essential in light energy utilization and electron transport chain for generation of ATP, which were all important components for photosynthesis of plants [55, 56]. Therefore, all these genes, which were involved in important process for plant growth and development, might evolve results of more frequent substitutions to adapt to different environmental conditions.

Among all positively selected genes, we found that the *accD* gene possessed the maximum number of sites under positive selection in Lardizabaloideae species, suggesting that the *accD* gene may play a pivotal role in the adaptive evolution of these species [57]. In addition, the likelihood ratio tests (LRTs) results showed that *p*-value of *rbcL* gene was less than 0.05, corroborating that sites in rubisco large subunit protein in clade Lardizabaloidae species have been under positive selection. As an important modulator of photosynthetic electron transport, recent study has revealed that positive selection of the *rbcL* gene was fairly common in all the main lineages of land plants [58, 59]. Thus, the *rbcL* gene was widely used to establish the diverse phylogenetic relationships of land plants [18, 60]. In summary, positive selection would possibly contribute to subfamily Lardizabaloideae diversification and adaptation.

The phylogenetic analysis in order Ranunculales
Chloroplast genome sequences which contained sufficient information have been widely used to reconstruct phylogenetic relationships among angiosperms even at lower taxonomic levels [61–64]. The phylogenetic relationships based on CCG dataset were consistent with the Angiosperm Phylogeny Group (APG) IV system of classification [19]. Unexpectedly, the phylogenetic relationships based on both CCG and concatenated protein-coding genes CDs datasets were inconsistent. The phylogenetic tree based on CDs dataset showed that the family Circaeasteraceae species were clustered into a clade with family Ranunculaceae species. This result indicates that species *Kingdonia uniflora* and *Circaeaster agrestis* in family Circaeasteraceae had strong support to be a sister to the *Pulsatilla chinensis* in family Ranunculaceae based on chloroplast protein-coding genes, which was inconsistent with the APG IV classification system. The inconsistent phylogenetic relationships implied a different rate of evolution in coding regions and non-coding regions, which might due to the nucleotide substitutions of non-coding regions were noisy than those.

Conclusions
This is the first report of the complete chloroplast genome sequence of *Stauntonia chinensis*. The architectural and the phylogenomic analysis of complete chloroplast genomes of eight Lardizabaloideae plants and relevant species could provide valuable genomic resource of this subfamily and its relatives. Meanwhile, several variation hotspots detected as highly variable regions could be served as the specific DNA barcodes. Our genomics analysis of these complete chloroplast genomes will lead to potential applications in the understanding of evolution and adaptation of species in the subfamily Lardizabaloideae.

Methods
Plant materials and DNA extraction
*Stauntonia chinensis*, which was identified by Prof. Liao Liang according to Flora of China, was sampled from Xianyan Mountain in Nanping city (118.10E, 26.73 N), Fujian Province, China. The voucher specimen deposited in Jiujiang University (accession number JU130801). Approximately 5 g of fresh leaves was harvested for genomic DNA isolation using an improved extraction method [65].

Chloroplast genome sequencing, assembly and annotation
A library with the insertion size of 430 bp was constructed, and all genome data were sequenced using an Illumina Hiseq 4000 platform at BIOZERON Co., Ltd. (Shanghai, China) [66]. The filtered reads were aligned with the *Akebia trifoliata* chloroplast genome (GenBank accession KU204898), and mapped to the reference chloroplast genomes [67, 68]. The chloroplast genes were annotated using an online DOGMA tool, using default parameters to predict protein-coding genes, transfer RNA (tRNA) genes, and ribosome RNA (rRNA) genes, coupled with manual check and adjustment [69].

Codon usage, and repeat structure
Codon usage was determined for all protein-coding genes using the program Codon W 1.44 [70]. The relative synonymous codon usage (RSCU) was calculated to examine the deviation in synonymous codon usage. Six values were used to estimate the extent of bias toward codons: the codon adaptation index (CAI), codon bias index (CBI), frequency of optimal codons (Fop), the effective number of codons (ENC), GC content (GC), and GC content of synonymous third codons positions (GC3s).

Repeat structures (forward, palindromic, complement, and reverse) within the chloroplast genomes were analyzed using REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer/), with following parameters: minimal repeat
size of 30 bp and hamming distance of three [71]. Tandem repeats were identified using the Tandem Repeats Finder 4.09 (http://tandem.bu.edu/trf/trf.html) with parameters being set as 2, 7, and 7 for alignment parameters match, mismatch, and indels, respectively [72]. The minimum alignments score and maximum period size were 50 and 500, respectively. Perl script MISA was used to determine single sequence repeats (SSRs) within these chloroplast genomes with parameters of mono-, di-, tri-, tetra-, penta-, and hexa-nucleotides being set as 10, 5, 4, 3, and 3, respectively [73].

**Genome comparison and nucleotide divergence**

Comparative chloroplast genomes of eight Lardizabaloideae species were carried out and visualized by using mVISTA online software (http://genome.lbl.gov/vista/index.shtml) [74] with the A. trifoliata as a reference [67]. The large structural changes such as gene order rearrangements, inversions, and insertions were identified using Mauve v2.4.0 with default settings [75]. The chloroplast genome borders were also analyzed to show the IR expansions and contractions. DNASeq v5.10 software was used to analyze the nucleotide diversity (Pi) and sequence polymorphism of Lardizabaloideae species [76].

**Species pairwise Ka/Ks ratios and positive selected analyses**

The concatenated single-copy gene coding sequences (CDs) of all 39 taxa were extracted and aligned with ClustalW [77]. Pairwise Ka/Ks ratios of all species were calculated using KaKs Calculator v2.0 [78]. The positive selected analyses were performed by an optimized branch-site model and Bayesian Empirical Bayes (BEB) method [53, 54]. The single-copy CDs of protein-coding genes of all 39 taxa were extracted and their amino acid sequences were aligned with ClustalW. The branch-site model was performed to test for potential positive selection using the CODEML algorithm implemented in EasyCodeML [79, 80]. The ratio (ω) of nonsynonymous to synonymous substitution rates was used to determine the selective pressure. The positive selection, no selection and negative selection were indicated when the ratio ω > 1, ω = 1, and ω < 1, respectively [80–82]. The likelihood-ratio tests (LRT) were performed according to Lan et al. [83]. The BEB method was used to compute the posterior probabilities of amino acid residues to identify whether these residue sites had potentially evolved under selection [53].

**Phylogenetic analyses**

The complete chloroplast genome (CCG) sequences and concatenated single-copy protein coding genes CDs of all 39 taxa were aligned using ClustalW. The phylogenetic analyses were carried out through maximum likelihood (ML) and Bayesian inference (BI) performed in IQ-TREE v1.6.1 and MrBayes 3.1.2, respectively [84, 85]. The best-fit models for both datasets were selected by MrModeltest v2.3. The Maximum likelihood analyses were conducted using IQ-TREE with 1000 bootstrap replicates. The BI analysis was run for 100,000 generations and sampled every 100 generations. The first 25% of the trees were discarded as burn-in, and the remaining trees were used to build a 50% majority-rule consensus tree.

**Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07484-7.

**Additional file 1.** The statistics of codon usage bias in all 39 taxa used in this study.

**Additional file 2.** Plastome alignment of all 39 taxa in this study. Gene arrangement map was carried out with only one copy of the IR using Mauve v2.4.0 software. The Akebia trifoliata chloroplast genome is shown at top as the reference genome. Local collinear blocks are represented by blocks of the same color connected by lines.

**Additional file 3.** Summary of Pairwise Ka/Ks ratios in Lardizabaloideae and other families.

**Additional file 4.** Partial alignment of amino acids sequences in the other 19 positively selected genes.

**Additional file 5.** Summary of complete chloroplast genomes of all 39 taxa in this study.

**Additional file 6.** The complete cp genome of Stauntonia chinensis.

Acknowledgements

We especially appreciate the conversation with the members of our group in developing some of the ideas presented in this study.

Authors’ contributions

FW designed the experiments and organized the manuscript. FW, and XW wrote the manuscript. FW, XW, TL, MJ, and XL performed the analyses. FW contributed to the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by National Natural Science Foundation of China [31560075, 31760047, 31960041], Natural Science Foundation of Jiangxi Province [20202BAB203045], and Foundation of Chinese medicine research of health and family planning commission of Jiangxi province [2017B070].

Availability of data and materials

All data generated or analyzed during this study were included in this published article and the Additional files. The complete cp genome of Stauntonia chinensis was submitted to GenBank under the accession number MAN401678, which could also be found in Additional file 6. All raw reads are available in the short sequence archive under accession no. PRJNA700993. All of the complete genome sequences used in this study were downloaded from NCBI (https://www.ncbi.nlm.nih.gov), and the accession numbers can be found in Additional file 5.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.
Competing interests
The authors declare that they have no competing interests.

Author details
1School of Pharmacy and Life Science, Jiujiang University, Jiujiang, China.
2State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, College of Plant Protection, Fujian Agriculture and Forestry University, Fuzhou, China.

Received: 28 September 2020 Accepted: 25 February 2021
Published online: 06 March 2021

References
1. Rogerio AP, Carlo T, Ambrosio SR. Bioactive natural molecules and traditional herbal medicine in the treatment of airways diseases. Evid Based Complement Alternat Med. 2016;2016:9872302.
2. Chen SL, Yu H, Luo HM, Wu Q, Li CF, Steinmetz A. Conservation and sustainable use of medicinal plants: problems, progress, and prospects. Chin Med. 2016;11:37.
3. Goron TL, Raizada MN. Genetic diversity and genomic resources available for the small millet crops to accelerate a new green revolution. Front Plant Sci. 2015;6:157.
4. Bendich AJ. Circular chloroplast chromosomes: the grand illusion. Plant Cell. 2004;16(7):1661–6.
5. Dong W, Liu J, Zhou J, Gao B, Xu C, Yao X, Zheng J, Gao J, et al. Complete plastome of Panax schinseng (Apostasioideae) reveals different evolutionary dynamics of IR/SSC boundary. BMC Evol Biol. 2018;19(7):1847.
6. Dong W, Xu C, Cheng T, Zhou S. Complete chloroplast genome of *Sedum sarmentosum* and *S. rupestre* in Saxifragales. PLoS One. 2013;8(10):e79665.
7. Asaf S, Khan AL, Khan AR, Waqas M, Khand M, Khan MA, Lee SM, Lee IJ. Analysis of plastome markers for species discrimination of *Sedum sarmentosum* and *S. rupestre* in the family Crassulaceae. Mol Genet Genomics. 2018;19(1):977–84.
8. Xie DF, Yu Y, Deng Y, Li J, Liu HY, Zhou SD, He XJ. Comparative analysis of the chloroplast genome of *Nicotiana otophora* and its comparison with related species. Front Plant Sci. 2016;7:843.
9. Dong W, Liu J, Yu J, Wang L, Zhou S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One. 2012;7(4):e35071.
10. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J, Zhou S et al. Five new bidesmoside triterpenoid saponins from *Stauntonia noroleaname*-type triterpenoid saponins from the stems of *Stauntonia yunnanensis* and *Machilus botomine* in the family Lauraceae. Front Plant Sci. 2015;6:662.
11. Saarela JM, Burke SV, Wysocki WP, Barrett MD, Craine JM, Peterson PM, Soreng RJ, Vorontsova MS, Duval MR. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. Peer J 2018;6:e4999.
12. Cho KS, Yoon BK, Yoon YH, Hong SY, Mekapouki M, Kim KH, Yang TJ. Complete chloroplast genome sequence of tartary buckwheat (*Fagopyrum tataricum*) and comparative analysis with common buckwheat (*F. esculentum*). PLoS One. 2015;10(5):e0125332.
13. Caron H, Dumas S, Marque G, Messier C, Bandou E, Petit RJ, Kreamer A. Spatial and temporal distribution of chloroplast DNA polymorphism in a tropical tree species. Mol Ecol. 2000;9(8):1089–98.
14. Ingvason PK, Ribstein S, Taylor DR. Molecular evolution of insertions and deletions in the chloroplast genome of silene. Mol Biol Evol. 2003;20(11):1737–40.
15. Park S, Ruhlman TA, Weng ML, Hajra IH, Sabir J, Janssen RK. Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of geranium. Genome Biol Evol. 2017;9(6):1766–80.
16. Ivanova Z, Sankov G, Daskalova E, Zhananova G, Apostolova E, Yahubyan G, Baev V. Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. Front Plant Sci. 2017;8:204.
rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol. 2014;31(3):645–59.
44. Skuzu L, Szuco I, Filp E, Strzala T. Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. PLoS One. 2019;14(2): e213023.
45. Yi DK, Lee HL, Sun BY, Chung MY, Kim KJ. The complete chloroplast DNA sequence of Elaeutheroctus senticosus (Araliaceae): comparative evolutionary analyses with other three astereids. Mol Cells. 2012;33(3):497–508.
46. Downie SR, Jansen RK. 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(1):336–51.
47. Rahemi A, Fatehiah E, Tahlai A, Hassan R, Gradziel T, Holta K, Chaparro J. Genetic diversity of some wild almonds and related Prunus species revealed by SSR and EST-SSR molecular markers. Plant Syst Evol. 2012;286(1):173–92.
48. Kumar M, Choi J, Kumari N, Pareek A, Kim S. Molecular breeding in Brassica for salt tolerance: importance of microsatellite (SSR) markers for molecular breeding in Brassica. Front Plant Sci. 2015;6:688.
49. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. Annu Rev Genomics Hum Genet. 2003;4:213–35.
50. Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J Mol Evol. 2000;50(1):56–68.
51. Tiffin P, Hahn MW. Coding sequence divergence between two closely related plant species: Arabidopsis thaliana and Brassica napus sp. ssp. pekinensis. J Mol Evol. 2002;54(6):746–53.
52. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
53. Yang Z, Wong WS, Nielsen R. Bayesian estimation of amino acid protein evolution. Annu Rev Genomics Hum Genet. 2003;4:213–35.
54. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486.
55. Yamori W, Shikanai T. Physiological functions of cyclic electron transport around photosystem I in sustaining photosynthesis and plant growth. Annu Rev Plant Biol. 2015;66:81–106.
56. Peltier G, Aro EM, Shikanai T. NDH-1 and NDH-2 plastoquinone reductases in oxygenic photosynthesis. Annu Rev Plant Biol. 2016;67:55–80.
57. Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, Li ZH. Molecular evolution of chloroplast genomes of Orchid species: insights into phylogenetic relationship and adaptive evolution. Int J Mol Sci. 2018;19(3):716.
58. Kapralov MV, Filatov DA. Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evol Biol. 2007;7:73.
59. Allahverdyyeva T, Mamedov F, Maempa P, Voss I, Aro EM. Modulation of photosynthetic electron transport in the absence of terminal electron acceptors: characterization of the rbcL deletion mutant of tobacco. Biochim Biophys Acta. 2005;1709(1):69–83.
60. Korall P, Kenrick P. Phylogenetic relationships in Selaginellaceae based on RBCL sequences. Am J Bot. 2002;89(3):506–17.
61. Carbonell-Caballero J, Abrao R, Ibarz V, Terol J, Talon M, Dopazo J. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus Citrus. Mol Biol Evol. 2015;32(3):105–17.
62. Sun L, Fang L, Zhang Z, Chang X, Penny D, Zhong B. Chloroplast phylogenetic inference of green algae relationships. Sci Rep. 2016;6:20528.
63. Zhang X, Zhou T, Kanwal N, Zhao Y, Bai G, Zhao G. Completion of eight Gynostemmmia BL. (Cucurbitaceae) chloroplast genomes: characterization, comparative analysis, and phylogenetic relationships. Front Plant Sci. 2017;8:1583.
64. Zhao ML, Song Y, Ni J, Yao X, Tan YH, Xu ZF. Comparative chloroplast genomics and phylogenetics of nine Lindera species (Lauraceae). Sci Rep. 2018;8(1):18844.
65. McPherson H, van der Merwe M, Delaney SK, Edwards MA, Henry RJ, McIntosh E, Rymer WD, Milner ML, Siow J, Rossetto M. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. BMC Ecol. 2013;13:8.
66. Borgstrom E, Lundin S, Lundeberg J. Large scale library generation for high throughput sequencing. PLoS One. 2011;6(4):e19119.