Molecular Biology Reports (2022) 49:3085–3098
https://doi.org/10.1007/s11033-022-07139-0

ORIGINAL ARTICLE

Complete chloroplast genome of *Campsis grandiflora* (Thunb.) schum and systematic and comparative analysis within the family Bignoniaceae

Haimei Chen[1] · Zhuoer Chen[2] · Qing Du[3] · Mei Jiang[1] · Bin Wang[2] · Chang Liu[1]

Received: 7 October 2021 / Accepted: 12 January 2022 / Published online: 21 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Background Plants belonging to the Bignoniaceae family have a wide distribution in the tropics and large populations around the world. However, limited information is available about Bignoniaceae. This study aimed to obtain more research information about Bignoniaceae plants and provide data support for the study of plant plastid genomes.

Methods and results In the present study, we focused on the chloroplast genome bio-information of *Campsis grandiflora.* The chloroplast DNA of *C. grandiflora* was extracted, sequenced, assembled, and annotated with corresponding software. Results show that the complete chloroplast genome of *C. grandiflora* is 154,303 bp in length and has a quadripartite structure with large single copy of 85,064 bp and a small single copy of 18,009 bp separated by inverted repeats of 25,615 bp. A total of 110 genes in *C. grandiflora* comprised 79 protein-coding genes, 27 transfer RNA genes, and 4 ribosomal RNA genes. The distribution of simple sequence repeats and long repeat sequences was determined. We carried out phylogenetic analysis based on homologous amino acid sequence among 45 species derived from Bignoniaceae. Compared with the chloroplast genome of *A. thaliana*, an inversion was identified in that of *C. grandiflora*, which result in the incomplete *clpP* gene.

Conclusions The chloroplast genomes were used for molecular marker, species identification, and phylogenetic studies. The outcome strongly supported that *C. grandiflora* and genus *Incarvillea* formed a cluster within Bignoniaceae. This study identified the unique characteristics of the *C. grandiflora* cp. genome, thus providing theoretical basis for species identification and biological research.

Keywords *Campsis grandiflora* · Chloroplast genome · Bignoniaceae · Gene rearrangement · *clp*P gene

Introduction

The chloroplast genome plays an important role in plant plastid genetic system, and its highly conserved circular quadripartite double-stranded structure consists of a large

Haimei Chen and Zhuoer Chen have contributed equally to this work.

✉ Chang Liu
cliu6688@yahoo.com

Haimei Chen
hmchen@implad.ac.cn

Zhuoer Chen
cze982920577@163.com

Qing Du
bjdfydq@163.com

Mei Jiang
mjiang0502@163.com

Bin Wang
beinwang@126.com

[1] Key Laboratory of Bioactive Substances and Resource Utilization of Chinese Herbal Medicine from Ministry of Education, Engineering Research Center of Chinese Medicine Resources from Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, 100193 Beijing, People's Republic of China

[2] School of Pharmacy, Xiangnan University, 423000 Chenzhou, Hunan, China

[3] College of Pharmacy, Qinghai Minzu University, 810007 Xining, Qinghai, China

(LSC; 80−90 kb) and small single-copy regions (SSC; 16−27 kb), separated by two inverted repeat regions (IRs) with length of 20–28 kb. This configuration leads to its low mutation rate during plant evolution. Therefore, the stable gene content, simple structure, non-recombinant, and mostly maternally inherited properties indicate that the chloroplast genomes contain a great deal of valuable biological information as an ideal material to support phylogeny and evolution studies [1]. With the rapid development of high-throughput sequencing technology in recent years, researchers have efficiently extracted and sequenced chloroplast genomes from plants, thus greatly advancing the process of chloroplast genome sequencing. Chloroplast genome sequencing information has been widely used to build the basis of phylogenetic analysis, and the evolutionary history of many plant groups has been deeply explored and supported [2].

The abundance of species in Bignoniaceae includes a total of 650 species in 120 genera, including *Catalpa*, *Campsis*, *Adenomocalymma*, *Amphilophium*, and *Anemopaegma* [3]. Bignoniaceae plants, which mainly include trees, shrubs, or woody vines, are widely distributed in the tropics and subtropics and are important tropical plants. The vast majority of species of Bignoniaceae have very large and beautiful flowers and various exotic fruit shapes and are cultivated in botanical gardens around the world, as ornamental, scenic, and street trees, and as an ideal shade pergola plant for the tropics [4]. *Campsis grandiflora* is a climbing vine affiliated with the genus *Campsis*, family Bignoniaceae. Distinguished from *Campsis radicans*, other plants of the same genus derived from North America, *C. grandiflora* is mainly distributed in China and Japan and cultivated in Vietnam, India, and Pakistan [5]. *C. grandiflora* can be used for ornamental and medicinal purposes. Pharmacological studies have shown that it has antibacterial, antithrombotic, and antitumor effects [6]. According to the Chinese Pharmacopoeia (2015 Edition) [7], *C. grandiflora* promotes blood circulation, and its flower is a diuretic for meridional treatment and can cure the disease of falling and injury [8].

Although the family Bignoniaceae has numerous species, only more than 40 chloroplast data have been recorded [9]. The chloroplast genome study of the entire genus *Campsis*, an important branch of Bignoniaceae, is still blank. In the present study, we obtained the chloroplast genome sequence of the *C. grandiflora* by using high throughput sequencing technology, characterized the gene contents, gene loss, IR border, genome rearrangements within the family Bignoniaceae, obtained phylogenetic information about *C. grandiflora* and its closely related species within the family Bignoniaceae. In summary, results obtained in this study provided valuable information to elucidate the evolutionary history of species in Bignoniaceae.

Materials and methods

Plant material, DNA purification, and genome sequencing

The *C. grandiflora* sample was collected in Huazhong Medicinal Botanical Garden, China (located at 109.76 E, 30.18 N) with voucher sample ID of implad201808016 (IMPLAD, China). The whole-genome DNA of *C. grandiflora* was extracted using the plant genomic DNA kit (Tiangen Biotech, Beijing, China). Library construction and genome sequence were completed using the Hiseq 2500 platform (Illumina, San Diego, CA, USA) [10].

Chloroplast genome assembly and annotation

The raw data of the sequence were assembled into a complete chloroplast genome with NOVOplasty (ver. 4.0.1) [11].

Genome annotation and repeat analysis were conducted using CPGAVAS2, DB 2 [12]. For the annotation of tRNA genes, both tRNAscan and ARAGORN were used to predict tRNA genes initially. Those prediction results from tRNAscan-SE for genes without intron are saved, while those prediction results from ARAGORN for genes with intron are saved. These saved tRNA genes were used to search tRNAdb based on sequence similarity (http://trna.bioinf.uni-leipzig.de/DataOutput/Search). According to the results of BLAST search, we determined the name of the tRNA gene as the best hits. As a result, the *trn*E-UUC, *trn*S-CGA and *trn*M-CAU were curated as *trn*I-GAU, *trn*G-UCC and *trn*I-CAU, respectively.

Phylogenetic analysis

To determine the phylogenetic position of *C. grandiflora* in Bignoniaceae, we used the maximum likelihood method [13] to construct an evolutionary tree with the cpREV model of IQ-Tree [14] for 56 common protein sequences of 45 species, including genus *Adenocalymma* [15], *Neojobertia* [16], *Pleonotoma* [16], *Amphilophium* [17], *Anemopaegma* [18], *Tanaecium* [19], *Dolichandra* [20], *Oroxylum* [21], *Catalpa* [22, 23], *Incarvillea* [24–26], *Spathodea* [27], and two outer groups (*Paulownia tomentosa* [28] and *Arabidopsis thaliana* [29]) of species from the family Bignoniaceae. For phylogenic tree construction, we used Phylosuite (version 1.2.2) [30] to extract the GenBank files of 47 species to obtain the common protein-coding genes sequences. Then, we conducted multiple sequence alignment of the common protein-coding genes by using MAFFT (v7.313). The MAFFT outcome of the common protein-coding genes was concatenated and the conserved blocks from multiple alignments were

calculated by Gblocks (v0.91b) for phylogenetic analysis. After we obtained the contree file, the visual work of the evolutionary tree was performed using iTOL Interactive Tree of Life [31].

Simple sequence repeat (SSR) and repeat analysis

The SSR locus and distribution were identified using the MIcroSAtellite identification tool [32]. The long tandem repeats (matching parameter = 2, mismatching and indel parameter = 7, minimum identity score = 50, maximum repeat period = 500, minimum repeat size = 30 bp, repeat unit similarity ≥ 90%) were identified using the tandem repeat finder [33]. The long interspersed repeats (repetition length ≥ 30 bp, Hamming distance = 3) were identified using the Vmatch (large scale sequence analysis software) [34].

Synteny analysis

In this study, we compared 45 Bignoniaceae species with *A. thaliana* to perform gene scale dot-plot analysis with Gepard (ver. 1.40 final) [35].

Genome rearrangements were identified between the chloroplast genome of *A. thaliana* and those of *A. oligoneuron* (NC_037232.1), *A. gnaphalanthum* (NC_042903.1), *T. tetragonolobum* (NC_027955.1), *A. paniculatum* (NC_042918.1), *I. compacta* (NC_050666.1), *I. sinensis* (NC_051523.1), *N. candolleana* (NC_036503.1), *A. allamandiflorum* (NC_036494.1), *A. biternatum* (NC_036496.1), *A. marginatum* (NC_037457.1), and *C. grandiflora* (MW430049), using BLASTN with an E-value cutoff of 1e-10. The homologous regions and gene annotations were visualized using a web-based genome synteny viewer Easyfig (ver. win2.1) [36].

Junction sites analysis

We used the GenBank files of 11 representative species with genomic structural variations from 45 species of Bignoniaceae that were used for detailed analysis to obtain the gene distribution on LSC, SSC, IRa, and IRb border. The location of genes on the boundaries was visualized using IRSCOPE [37].

Non-synonymous replacement (Ka)/synonymous replacement (Ks) analysis

We used the (adaptive branch-site random effects likelihood) aBSREL model of Hyphy Vision software to contribute the selective pressure analysis [38] among 45 species in Bignoniaceae. We first acquired the corresponding chloroplast genome GB and FASTA files according to the accession number in NCBI. Then, 63 clusters of orthologous genes were obtained among these species to calculate the Ka/Ks. The outcome was listed in aBSREL. json format. In the present study, we selected genes with p value < 0.05. The detailed information is shown in the web version of aBSREL.

Results

Genome organization and compositions

The chloroplast genome sequence (GenBank accession no.: MW430049) of *C. grandiflora* was a typical circular DNA molecule with a total length of 154,303 bp. It has a conservative tetrad structure consisting of an LSC region, an SSC region, and a pair of IR regions, with lengths of 85,064, 18,009, and 25,615 bp, respectively (Fig. 1). The G/C content of the chloroplast genome of *C. grandiflora* was 38.09%. The G/C content in the IR region (43.17%) was higher than that in the SSC (32.74%) and LSC regions (36.16%).

Gene content

The chloroplast genome of *C. grandiflora* encodes 110 unique genes, including 79 protein-coding genes, 27 transfer RNA (tRNA) coding genes, and 4 ribosome RNA (rRNA) coding genes (Table S1). Among these genes, eight protein coding genes (*rps*12, *ndh*B, *rpl*2, *rpl*23, *rps*7, *ycf*1, *ycf*2, and *ycf*15), 7 tRNA coding genes (*trn*A-UGC, *trn*E-UUC, *trn*L-CAA, *trn*M-CAU, *trn*N-GUU, *trn*R-ACG, and *trn*V-GAC) and 4 rRNA coding genes (*rrn*16S, *rrn*23S, *rrn*5S, and *rrn*4.5 S) were located in the IR region. Twelve protein-coding genes (*rps*16, *atp*F, *rpo*C1, *pet*B, *pet*D, *rpl*16, *rpl*2 (+), *rpl*2(−), *ndh*B(+), *ndh*B(−), and *ndh*A) contain one intron, and one protein-coding genes (*ycf*3) contain two introns. Eight tRNA coding genes (*trn*K-UUU, *trn*G-UUC, *trn*L-UAA, *trn*V-UAC, *trn*I-GAU(−), *trn*I-GAU(+), *trn*A-UGC (−), and *trn*A-UGC (+)) contain one intron (Table S2). We also found the *clp*P gene became a pseudogene, unable to encode a complete protein.

The coding sequence (CDS) in the chloroplast genome of *C. grandiflora* was 79,170 bp, accounting for 51.31% of the total genome length. The length of the rRNA genes was 9388 bp, accounting for 6.08% of the whole genome length. The length of the tRNA genes was 2811 bp, accounting for 1.82% of the whole genome length. The non-coding regions of the *C. grandiflora* chloroplast genome mainly includes introns and gene spacers, whose length accounts for 40.79% of the whole genome length.

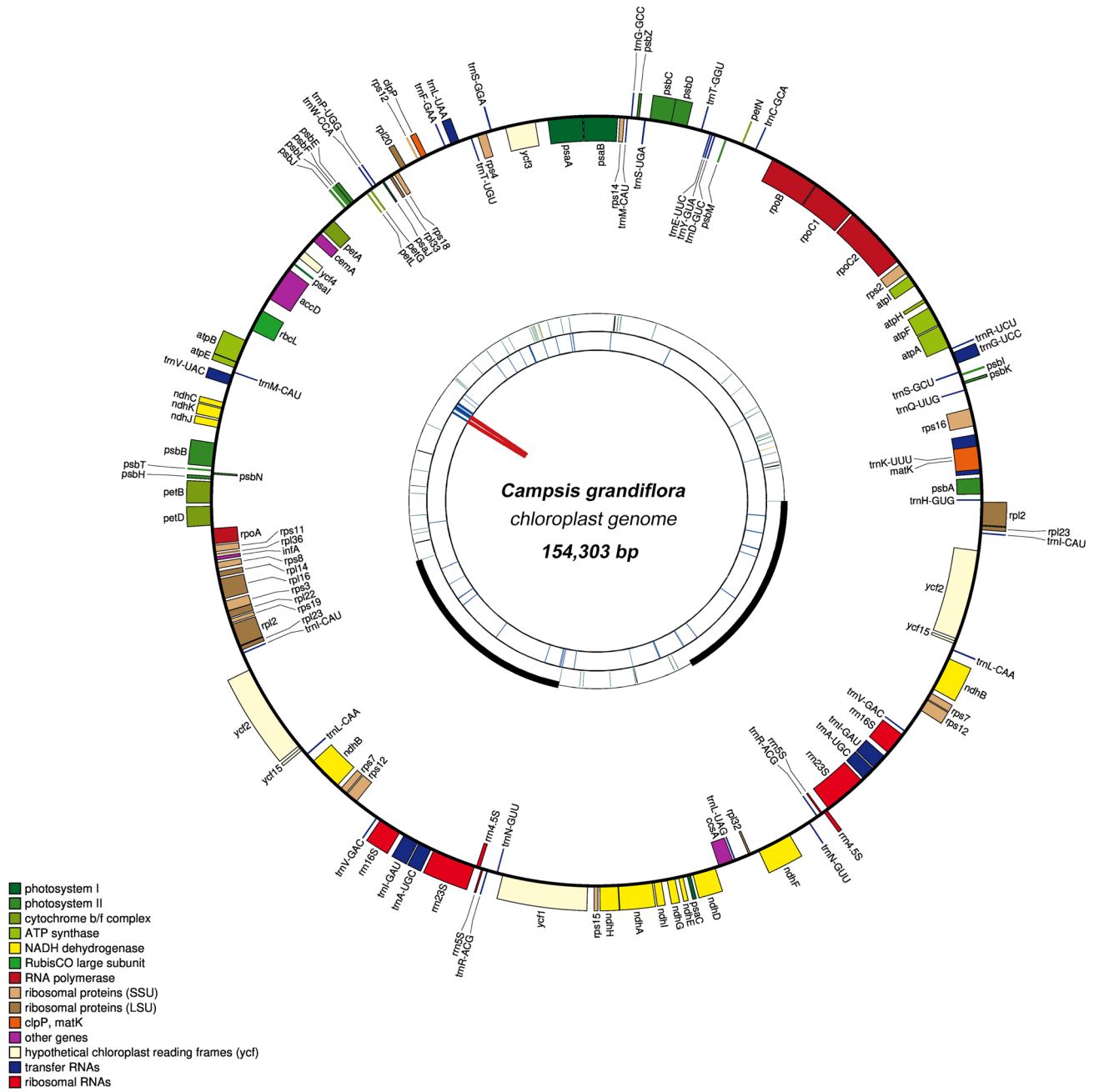

Fig. 1 Map of the chloroplast genome of *Campsis grandiflora*. Four rings are observed in the figure: from the center outwards, the red and green arcs in the first circle represent the forward and reverse repeating sequence, respectively. The short bars in the second circle represent tandem repeats. The short bar in the third circle represents the microsatellite repetition sequence. The fourth circle is the genetic structure and location map of the chloroplast genome. Genes with different functions are shown in different colors. (Color figure online)

SSR and repeat sequences analysis

The repeat sequences are particular nucleic characteristic sequence repeat units with multiple copies in the genome. These repeats might play a significant role in the evolution of the chloroplast genome and can be used for species identification and molecular breeding as molecular markers. The repeat sequences are classified into three forms, namely, SSR, long tandem repeats, and long interspersed repeated sequence according to their length and correlation [1].

SSR is also named microsatellite sequence. It is a piece of DNA that consists of multiple duplicate basic repeat units made of 1–6 nucleotides. The SSR is widespread all around the different places of the gene. Their length is usually below

200 bp. We analyzed and listed the quantity, type, size, and locus of SSRs in the chloroplast genome of *C. grandiflora*. In total, 59 SSRs were identified in the *C. grandiflora* chloroplast genome. These SSRs are mainly composed of mononucleotide and dinucleotide repeat units (Table S3). No other forms such as tri-, tetra-, penta-, and hexa-nucleotide repeat units were found. Most of the 59 SSRs we found in the intergenic spacers (35 SSRs), 9 SSRs were located in the coding sequences, and 7 SSRs were situated in the introns of particular genes (Table S4).

The long tandem repeats refer to the repeated repetition of a sequence on a chromosome. A total of 40 tandem repeats have been found, satisfying the two conditions that the total length is over 20 bp, and the similarity between repeating units is greater than or equal to 90% (Table S5). We also listed the related property in the table. Among the long tandem repeats, more than half (22) of the repeats were located in IGS, 16 repeats are shown in the CDS, the one remainder repeats were located in the intron of gene.

Interspersed repeats are another kind of repeated sequence different from tandem repeats. It includes palindromic and direct repeats. With the *e* value less than 1E-4 as the threshold, the scattered repeats of plumbic chloroplast genomes included 49 direct repeats. Notably, all of the interspersed repeats of *C. grandiflora* chloroplast genome are D type (direct repeat sequence). These interspersed repeats are all in the range of 62,500–63,700 of *acc*D gene, and almost all of them are located in the non-coding region, except for one sequence that its repeat unit I in the CDS of *acc*D (Table S6).

Phylogenetic analysis

To obtain the phylogenetic information of *C. grandiflora* and make valid hypotheses about the homology between different lineages of Bignoniaceae, we used 45 Bignoniaceae species and 2 outgroup species chloroplast genomes to construct the phylogenetic tree of Bignoniaceae (Fig. 2).

The tree shows that two primary branches initially diverged from the tree root. Fifteen species from the genus *Adenocalymma*, *Neojobertia*, and *Pleonotoma* gathered into a branch on the tree. Eleven species of genus *Amphilophium* converged into a branch. Eight species of genus *Anemopaegma* converged into a branch. Then, genus *Amphilophium*, *Anemopaegma*, *Tanaecium*, and *Dolichandra* gathered into a big branch with *Adenocalymma*, *Neojobertia*, and *Pleonotoma*. Furthermore, the grand branch congregated a branch with genus *Oroxylum*, and then the genus *Spathodea*. Two species of genus *Catalpa* gathered into a branch. From this view, the eight genera mentioned above have contributed to the upper grand branch of the evolutionary tree of the family Bignoniaceae. In the remaining part of the tree, three species of genus *Incarvillea* gathered into a branch, and then *Tecomaria* have aggregated a branch with genus *Incarvillea*. At last, genus *Campsis*, *Incarvillea*, and *Tecomaria* have converged into another grand branch of the tree. These results indicate that the closest sister genus of *Campsis* is *Incarvillea* and *Tecomaria* in Bignoniaceae,

In the phylogenetic tree of the family Bignoniaceae, the bootstrap scores of all branches of the evolutionary tree were high (≥ 47%), indicating that the evolutionary tree has high reliability. The results of the phylogenetic analysis are consistent.

Synteny analysis

To identify the genome rearrangement of Bignoniaceae, we selected the cp. genome sequences of *C. grandiflora* and other 44 species belonging to Bignoniaceae for synteny analyses (Table 1). These 44 species include *Adenocalymma* (13), *Anemopaegma* (8), *Amphilophium* (11), *Catalpa* (2), *Dolichandra* (1), *Oroxylum* (1), *Pleonotoma* (1), *Spathodea* (1), *Incarvillea* (3), *Tanaecium* (1), *Tecomaria* (1), *Neojobertia* (1), respectively (Table 1). According to whether the structure was inverted and whether the IR region was expanded, these genomes were classified into 10 types compared with *A. thaliana*. The first group includes *Anemopaegma acutifolium*, *Anemopaegma arvense*, *Anemopaegma glaucum*, *Anemopaegma foetidum*, *Anemopaegma album*, *Anemopaegma chamberlaynii*, *Anemopaegma prostratum*, *Anemopaegma oligoneuron* which are all belonged to *Anemopaegma*. There was an inversion in the LSC region of the chloroplast genomes of this group compared with that of *A. thaliana*, which results in the *ycf*2 gene being transcribed counterclockwise. Meanwhile, the IR region underwent expansion, resulting in the duplication of truncated *rps15*, ycf1, genes included ancestral angiosperm IR regions (*trnR, trnN, rrn5, rrn4.5, rrn23, trnA, trnI, rrn16, trnV, rps12, rps7, ndhB, trnL, ycf2, trnI, rpl23, rpl2*), *rps19*, *rpl22*, *rps3*, *rpl16*, *rpl14*, *rps8*, *infA*, *rpl36*, *rps11*, *rpoA*, *petD*, truncated *petB* in the IRs (Fig. 3A). The second group includes *Adenocalymma acutissimum, Adenocalymma trifoliatum, Adenocalymma aurantiacum, Adenocalymma bracteatum, Adenocalymma divaricatum, Adenocalymma peregrinum, Adenocalymma cristicalyx, Pleonotoma albiflora, Adenocalymma pedunculatum, Amphilophium gnaphalanthum, Amphilophium lactiflorum, Amphilophium chocoense, Amphilophium cuneifolium, Dolichandra cynanchoides, Oroxylum indicum, Spathodea campanulate, Catalpa bungee, Catalpa ovata* (Fig. 3B), whose chloroplast genome structure is similar to that of *A. thaliana*, except the duplication of truncated *rps15* and *ycf1* in the IR region. The second group contains the most species of Bignoniaceae. The third group includes *Amphilophium carolinae, Amphilophium dolichoides, Amphilophium steyermarkii, Amphilophium dusenianum, Amphilophium ecuadorense,*

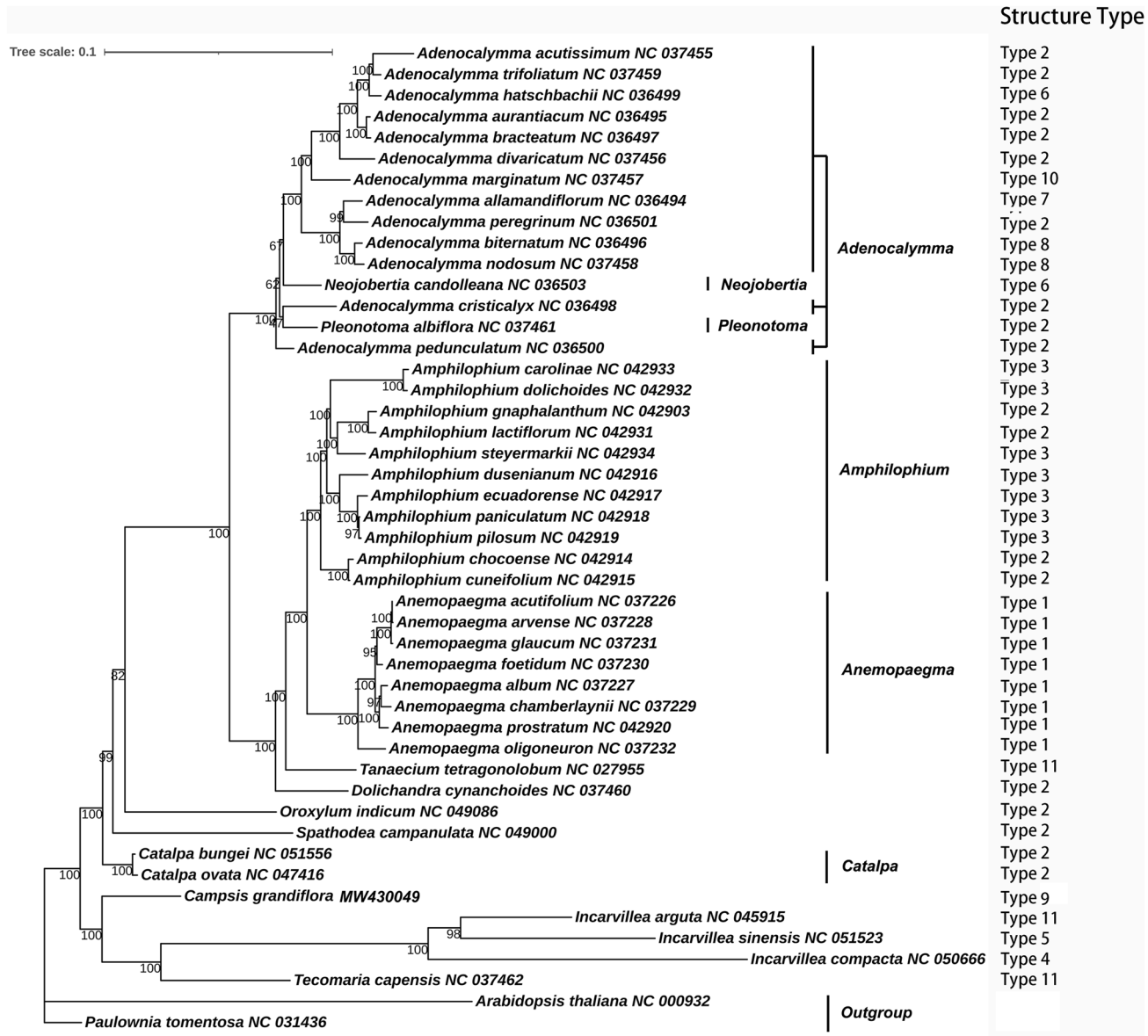

Fig. 2 Evolutionary tree of family *Bignoniaceae*. The phylogenetic results included 45 species within families and 2 outer species. The *N. candolleana* and *P. albiflora* interspersed in 13 *Adenocalymma* species converged into a large clade together with 11 *Amphilophium* species and 8 *Anemopaegma*, *T. tetragonolobum*, and *D. cynanchoides* species. This large clade subsequently converged with two species in the genus *Catalpa* and eventually gathered at the base of the evolutionary tree *with C. grandiflora*, three species in the genus *Incarvillea*, and *T. capensis*. According to the evolutionary tree, the event of *C. grandiflora* differentiation occurred in a relatively early period and has a close genetic relationship with *Incarvillea*. The right of the panel is the structure type of species according to the Fig. 4 and S3

Amphilophium paniculatum, Amphilophium pilosum whose chloroplast genome structure is similar to that of the first group but without the small inversion in the LSC region (Fig. 3C). The fourth group includes *Incarvillea compacta*, whose IR region contains truncated *rps15, ycf1, and* ancestral angiosperm IR region and also a large inversion in LSC region (Fig. 3D). The fifth group includes *Incarvillea sinensis, whose IR region contains ndhA, ndhH, rps15, ycf1, and* ancestral angiosperm IR region (Fig. 3E). The sixth group includes *Adenocalymma hatschbachii, Neojobertia candolleana* whose genome contains structural variation in the IR region (Fig. S2F). The seventh group includes *Adenocalymma allamandiflorum*, whose genome includes an inversion in LSC region (Fig. S2G). The eighth group includes *Adenocalymma biternatum, Adenocalymma nodosum*, whose genome contains an inversion in the LSC region (Fig. S2H). The tenth group includes *Adenocalymma marginatum* whose genome contains the 50 kb inversion in the LSC region (Fig.

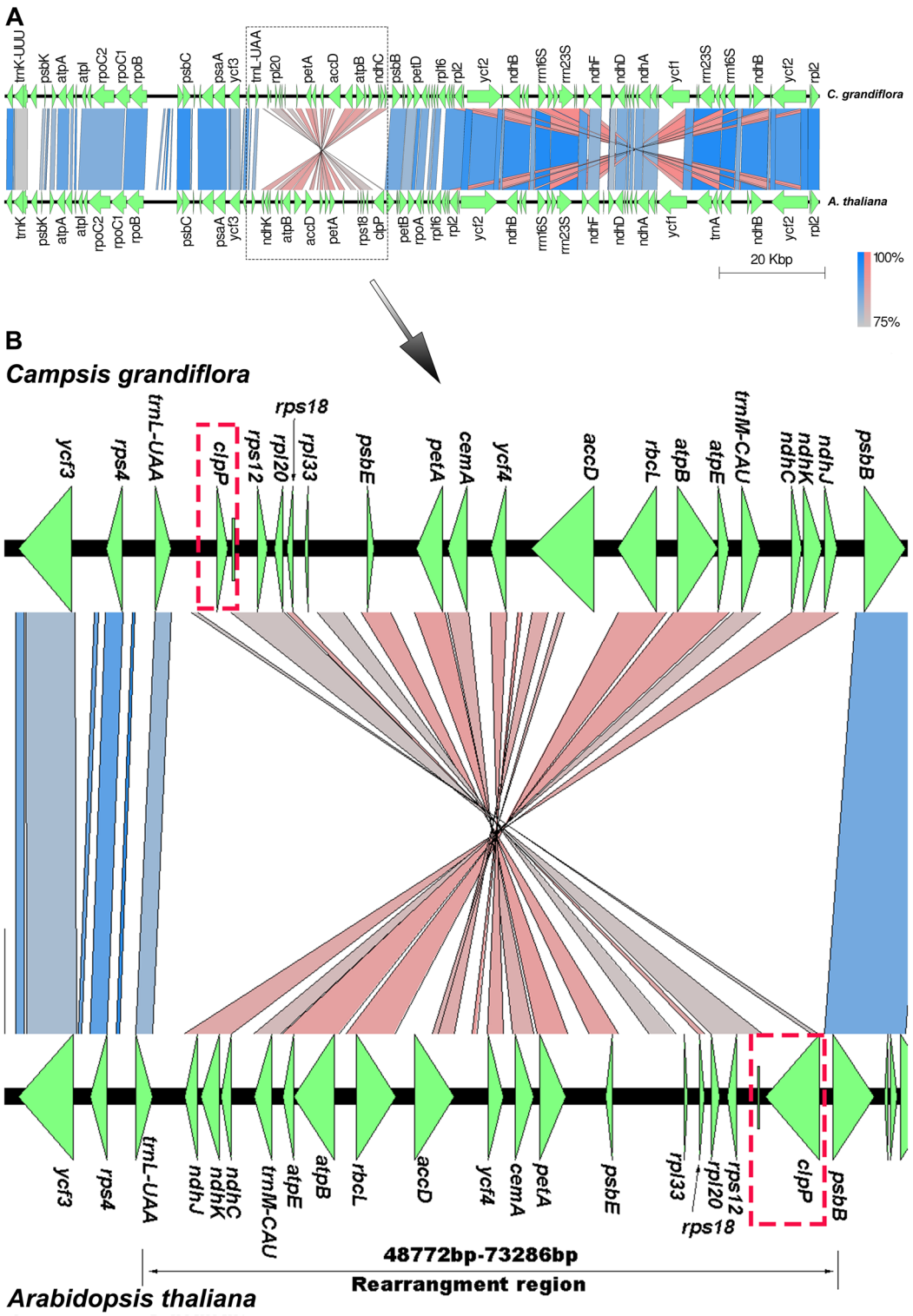

Fig. 3 Comparative genomic analyses of *A. thaliana* and five other representative species of the family Bignoniaceae. The chloroplast genome of *A. thaliana* was aligned with those of five species. Each horizontal black line represents one genome. The species names, accession numbers and the structure types are shown to the right of the corresponding line. The conserved regions are bridged by lines. Panels A to E show that the five types of genome structure respectively from *A. oligoneuron, A. gnaphalanthum, A. paniculatum, I. compacta, I. sinensis.* The blue and red bar represents the identity of forward and reverse comparison, respectively

S2I). The IR regions of seventh, eighth and tenth groups are similar to the IR region of the second group. The ninth group includes *C. grandiflora* whose genome contains an inversion in the LSC region (Fig. 4A). We next performed a genome comparison compared by using Gepard (ver. 1.40 final). The visualization result shows that the rearrangement occurred at 48,772–73,286 bp in the *C. grandiflora* chloroplast genome (Figure S1) which result in the incomplete of *clp*P gene (Fig. 4B). The eleven group includes *Tanaecium tetragonolobum, Incarvillea arguta, and Tecomaria capensis* whose genome contains no inversion (Fig. S2J). These results suggested that inversions frequently occurred in the evolution of Bignoniaceae.

.

Comparative analysis of gene loss in family Bignoniaceae

This study determined the correlation between gene loss and the rearrangement of genome structure. We made detailed statistics of the protein-coding gene loss in the particular plants of Bignoniaceae. All the plants involved in the statistics are derived from phylogenetic trees (Fig. 2). Based on the statistical results, The number of genes in the eight species from the genus *Anemopaegma* was highly conserved and consistent. In terms of gene loss, the *acc*D gene was lost in the genus *Incarvillea*. The *clp*P gene was found lost in *I. arguta* and *T. tetragonolobum* and had incomplete structure in *C. grandiflora*. The *ycf*15 gene was only found in *T. tetragonolobum, D cynanchoides, S campanulate, C. bungee, C. ovata, C. grandiflora, I compacta, T capensis.* In general, most of the gene loss occurred in the genus *Incarvillea and Tanaecium*.

Ka/Ks selective pressure analysis

In terms of genetics, Ka/Ks or dN/dS represents the ratio between non-synonymous replacement (Ka) and synonymous replacement (Ks). This ratio can be used to determine whether selective pressure acts on the protein-coding gene [39]. Nucleotide variations that do not lead to amino acid changes are called synonymous mutations, whereas non-synonymous mutations occur. Generally, synonymous mutations are not subject to natural selection, whereas non-synonymous mutations are. In evolutionary analysis, the rate at which synonymous and non-synonymous mutations occur should be determined [39].

In the present study, we used the phylogenetic tree (Fig. 2) as species reference and utilized the aBSREL model of software Hyphy for the selection pressure analysis of protein-coding genes (Table S7). Six genes were positively selected, including *ndh*G, *rbc*L, *rpl*22, *rpl*23, *rps*12, and *rps*15. In species *A. bracteatum*, the *ndh*G gene is positively selected. In species *A. glaucum* and *A. divaricatum*, the *rbc*L gene was positively selected. The *rpl*22 gene was positively selected in species *A. steyermarkii* and *D. cynanchoides*. In species *A. allamandiflorum* and *A. chamberlaynii*, *rpl*23 gene was positively selected. In *C. ovata*, *rps*12 and *rps*15 were positively selected. In species *C. grandiflora*, *rps*15 was positively selected.

IR expansion and contraction To unravel the gene distribution of junction site and compare the distinction between *C. grandiflora* and other species with genome rearrangement structure in the family Bignoniaceae, we visualized the gene distribution with IRSCOPE (Fig. S3).

Based on the result of visualization, the complete genome was divided into five parts with four vertical bars. The five parts include LSC, IRb, SSC, IRa, and LSC. Except for the *T. tetragonolobum, C. grandiflora, I. arguta, I. sinensis* and *T. capensis*, in the most species from Bignoniaceae, the *rps*15 gene has crossed the JSA between SSC and IRa (Table 1 and Fig. S3), In *I. sinensis*, the *ndh*F crossed the JSB between IRb and SSC regions. Notably, significant differences were observed in the length of SSC and IR regions between the species from genus *Incarvillea*. The SSC region in *I. sinensis* was only 8,666 bp in length, and the IR regions was 35,394 bp in length. However, in *I. compacta*, the SSC region reached a length of 21,925 bp. The length of genomic regions also differed in the genus *Amphilophium*. In the third structure type species, the gene that crossed the IRb and LSC are *pet*D. While in the species of *A. chocoense* and *A. cuneifolium*, the counterpart gene is *rpl2*. The expansion and contraction of the IR region led to the difference in IR length. For example, *A. paniculatum* and *A. oligoneuron*, their IR regions have reached 37,372 and 39,614 bp, which are much longer than the of *A. cuneifolium* with 27,814 bp. In the genus of *Anemopaegma*, the expansion of IR region to *pet*B gene leads to the longest length of IR region in order Lamiales. These results suggested that the contraction and expansion of IR region are consistent with evolution.

Discussion

In the current study, we extracted and sequenced the chloroplast genome of *C. grandiflora*. The raw data were assembled and annotated with relevant tools, and the complete information of the transiting chloroplast genome was obtained. Furthermore, phylogenetic analysis of *C. grandiflora* was performed. We obtained the rearrangement structure in the genome of *C. grandiflora* compared with that of *A. thaliana* (Fig. 4). The synteny analyses between species from the family Bignoniaceae and *A. thaliana* were also conducted. This information could provide us a new direction of chloroplast genome research of *C. grandiflora*.

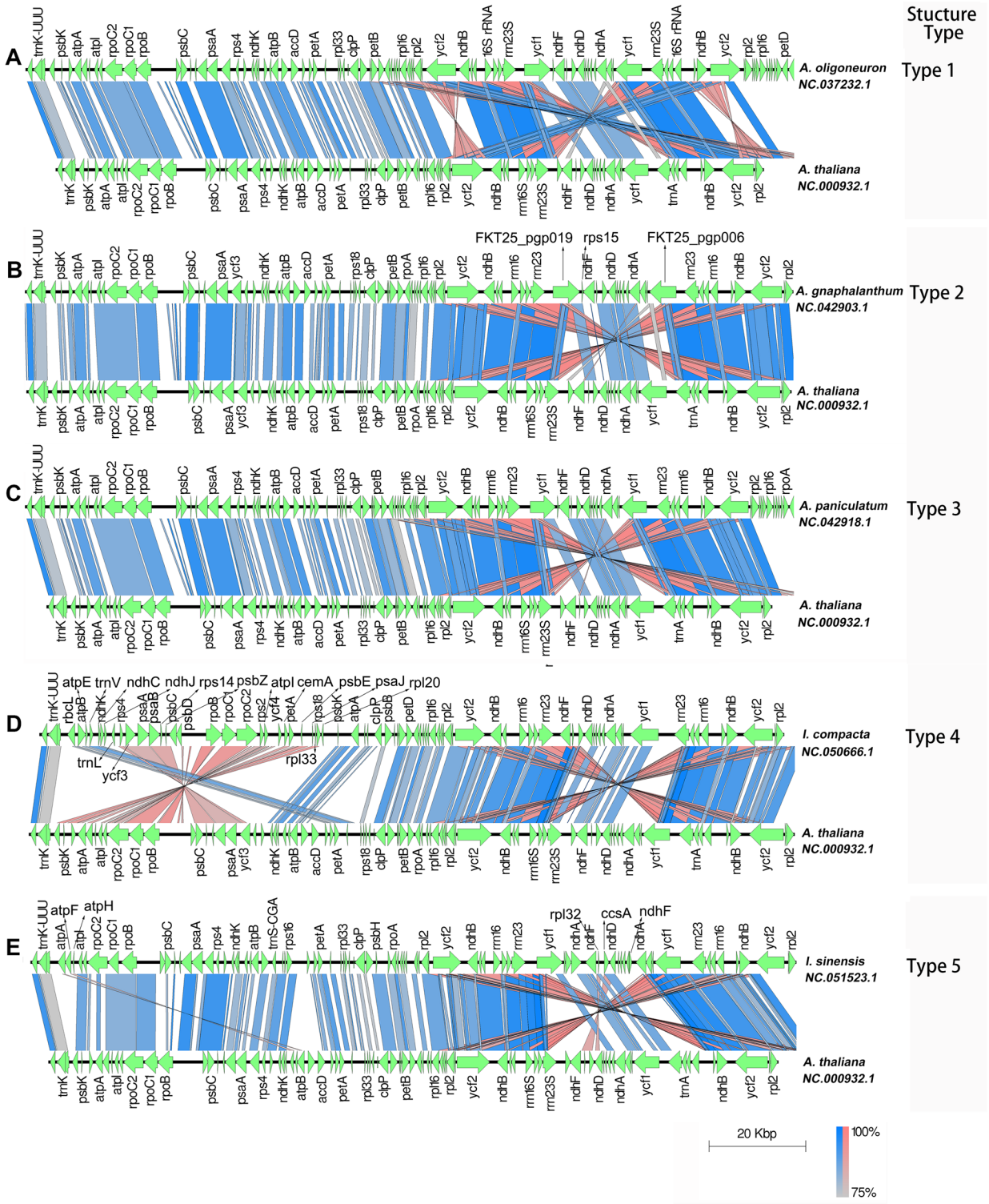

Fig. 4 Synteny analysis of *C. grandiflora* and *A. thaliana.* A panel shows that each horizontal black line represents a genome. The species names are shown to the right of the corresponding line. The green arrows represent genes, and the direction of the arrows indicates where the genes start and end on the genome. In the alignment of the two sequences, the conserved regions are bridged by lines, and the matching genes in the same direction are connected by blue lines. The reverse and matching genes are connected by red lines. The darker the color, the better or t less the match. B panel shows the details of the genome rearrangement area of *C. grandiflora*

Table 1 Comparison of IR contents in the chloroplast genomes of the structurally variant species in family Bignoniaceae

Accession no.	Structure type	Name of species	*ndhF*	*ndhI*	*ndhA*	*ndhH*	*rps15*	*ycf1*	*trnR, trnN, rrn5, rrn4.5, rrn23, trnA, trnI, rrn16, trnV, rps12, rps7, ndhB, trnL, ycf2, trnI, rpl23, rpl2*	*rps19*	*rpl22*	*rps3*	*rpl16*	*rpl14*	*rps8*	*infA*	*rpl36*	*rps11*	*rpoA*	*petD*	*petB*
NC_037455.1	Type 2	*A. acutissimum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037459.1	Type 2	*A. trifoliatum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036499.1	Type 6	*A. hatschbachii*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036495.1	Type 2	*A. aurantiacum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036497.1	Type 2	*A. bracteatum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037456.1	Type 2	*A. divaricatum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037457.1	Type 10	*A. marginatum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036494.1	Type 7	*A. allamandiflorum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036501.1	Type 2	*A. peregrinum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036496.1	Type 8	*A. biternatum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037458.1	Type 8	*A. nodosum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036503.1	Type 6	*N. candolleana*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036498.1	Type 2	*A. cristicalyx*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037461.1	Type 2	*P. albiflora*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_036500.1	Type 2	*A. pedunculatum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_042933.1	Type 3	*A. carolinae*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	1	1
NC_042932.1	Type 3	*A. dolichoides*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	1	1
NC_042903.1	Type 2	*A. gnaphalanthum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_042931.1	Type 2	*A. lactiflorum*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_042934.1	Type 3	*A. steyermarkii*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	1
NC_042916.1	Type 3	*A. dusenianum*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	1
NC_042917.1	Type 3	*A. ecuadorense*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	1
NC_042918.1	Type 3	*A. paniculatum*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	1
NC_042919.1	Type 3	*A. pilosum*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	1
NC_042914.1	Type 2	*A. chocoense*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_042915.1	Type 2	*A. cuneifolium*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037226.1	Type 1	*A. acutifolium*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_037228.1	Type 1	*A. arvense*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_037231.1	Type 1	*A. glaucum*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t

Table 1 (continued)

Accession no.	Structure type	Name of species	*ndhF*	*ndhI*	*ndhA*	*ndhH*	*rps15*	*ycf1*	*trnR, trnN, rrn5, rrn4.5, rrn23, trnA, trnI, rrn16, trnV, rps12, rps7, ndhB, trnL, ycf2, trnI, rpl23, rpl2*	*rps19*	*rpl22*	*rps3*	*rpl16*	*rpl14*	*rps8*	*infA*	*rpl36*	*rps11*	*rpoA*	*petD*	*petB*
NC_037230.1	Type 1	*A. foetidum*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_037227.1	Type 1	*A. album*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_037229.1	Type 1	*A. chamber-laynii*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_042920.1	Type 1	*A. prostratum*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_037232.1	Type 1	*A. oligoneuron*	1	1	1	1	t	2	2	2	2	2	2	2	2	2	2	2	2	2	t
NC_027955.1	Type 11	*T. tetragonolobum*	1	1	1	1	1	t	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037460.1	Type 2	*D. cynanchoides*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_049086.1	Type 2	*O. indicum*	1	1	1	1	t	2	2	2	1	1	1	1	1	1	1	1	1	1	1
NC_049000.1	Type 2	*S. campanulata*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_051556.1	Type 2	*C. bungei*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_047416.1	Type 2	*C. ovata*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
MW430049	Type 9	*C. grandiflora*	1	1	1	1	1	t	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_045915.1	Type 11	*I. arguta*	1	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1
NC_051523.1	Type 5	*I. sinensis*	t	1	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_050666.1	Type 4	*I. compacta*	1	1	1	1	t	2	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_037462.1	Type 11	*T. capensis*	1	1	1	1	1	t	2	1	1	1	1	1	1	1	1	1	1	1	1
NC_000932.1	Type 2	*A. thaliana*	1	1	1	1	1	t	2	t	1	1	1	1	1	0	1	1	1	1	1
NC_031436.1	Type 2	*P. tomentosa*	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1	1

1 indicates gene at SSC region. 0.2 indicates: gene located in IR region. T indicates the gene partially duplicated in IR region. The ancestral angiosperm IR region included *trnR, trnN, rrn5, rrn4.5, rrn23, trnA, trnI, rrn16, trnV, rps12, rps7, ndhB, trnL, ycf2, trnI, rpl23*, and *rpl2* genes

Special distribution of interspersed repeated sequences in *acc*D gene

Based on the analysis of repeated sequences, we found the particularity of interspersed sequences. In comparison with other species in this family, the interspersed sequences in *C. grandiflora* chloroplast genome showed obvious centralization and uniformity. The results showed that all the interspersed repeated sequences were distributed in the coding region of *acc*D gene. The distribution range is concentrated in 62,000–64,000 bp. In addition, the types of repeated sequences are only direct sequences, and palindrome sequences are not found (Table S6).

The acetyl-CoA carboxylase (*acc*D) gene is present in plastids such as chloroplasts in most flowering plants, including non-photosynthetic parasites. Its function is to encode the β-carboxylase subunit of acetyl-CoA carboxylase, thereby participating in plant life activities and material metabolism. Previous studies on tobacco have shown that if the *acc*D gene is knocked out or destroyed and cannot be successfully expressed in plastids, the leaf development of the plant will be severely affected. The loss of tissue cells leads to the stagnation of leaf division and differentiation, causing the failure of photosynthesis and the death of plants. Therefore, the *acc*D gene is an essential gene in plants. In the present study, the special distribution of interspersed sequences raised the possibility of molecular markers for the unique sequence in the gene coding region, and based on the statistics and analysis of the location of different repeat sequence families in different genes, new interspecies relationships or evolutionary processes can be found. These new directions are expected to be realized in future research.

Phylogenetic tree

Based on the distribution of species displayed in the phylogenetic tree, the genus *Adenocalymma* has a distant genetic relationship with the genus *Campsis*. By contrast, the genus *Incarvillea*, *Tecomaria*, and *Catalpa* have a closer genetic relationship with the genus *Campsis*. Considering that *C. grandiflora* is located at the base of the whole tree, the divergence event occurred in an earlier period of the evolution process in Bignoniaceae.

IR expansion and contraction

The results showed that the location and species of boundary genes were different with the length of genome sequence (Fig. S3). Therefore, the variation in the length of genomic regions leads to differences in the genes located at the boundaries. In *C. grandiflora* and *T. tetragonolobum*, the *ycf*1 gene was located at JSB and JSA, whereas in *C. grandiflora*, *rps*19 was located at the LSC region but crossed the JLB in *T. tetragonolobum*.

Systematic analysis of genome rearrangement that occurred in Bignoniaceae

We verified whether other species in the Bignoniaceae underwent genome rearrangement. We then analyzed 44 other species from the phylogenetic tree with Gepard (ver. 1.40 final). Finally, we identified 11 genomic structures in chloroplast genomes from 45 species of Bignoniaceae. We used EasyFig to visualize these 11 rearrangement structures (Fig. 3 and Fig. S2). Eight species of the genus *Anemopaegma* share the same genomic rearrangement [18]. In combination with the above-mentioned statistical results of gene content of IR region and the results of synteny analysis (Fig. 3, Fig. S3 and Table 1), in genus *Anemopaegma*, 8 species had the same genome structure and maintained a highly conservative gene number. This property can be considered as an intergeneric characteristic of the genus *Anemopaegma*. The second structure type contains the most species of Bignoniaceae. Meanwhile the IR regions of the seventh, eighth and tenth groups are similar to that of the second group. It is proposed that the second type structure is located at the base node of evolution. The chloroplast genome of *C. grandiflora* contains an inversion in the LSC region (Fig. 4A). The rearrangement occurred at 48,772–73,286 bp in the C. grandiflora chloroplast genome (Fig. S1) which result in the incomplete of *clp*P gene (Fig. 4B). Among species from the genus *Incarvillea*, the gene content of IR region was also significantly different, which displays rapid variation in the genus.

Conclusions

In the present study, we extracted, assembled, sequenced, and annotated the complete chloroplast genome of *C. grandiflora*, filling in the gaps in chloroplast genome information of genus *Campsis*. The phylogenetic analysis reveals the phylogenetic information of Bignoniaceae as well as the overall evolutionary history of 45 species of the family. The repeat sequence analysis also revealed the genetic characteristic information. The Ka/Ks analysis indicated the direction of evolution of Bignoniaceae. We conducted a detailed and in-depth analysis of the chloroplast genome of *C. grandiflora* and found that the chloroplast genome has an inverted rearrangement structure through synteny analysis. We also found and sorted out the rearrangement structures of 11 chloroplast genomes of Bignoniaceae from the available data by synteny analysis. The results will provide important phylogenetic information of *C. grandiflora*. Gene loss analysis was used to

determine the relationship between rearrangement structure and the gene quantity variation.

The Bignoniaceae family includes many species, but limited information is currently available. The results of this study are based on all the released chloroplast genome sequences available so far. With the acceleration of sequencing progress, the database of Bignoniaceae will be enriched day by day in the future, and more information will be discovered.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-022-07139-0.

Acknowledgements We would like to thank Prof. You Jinwen for identifying the plant materials.

Author contribution CL and HMC conceived the study. MJ collected the samples of *C. grandiflora*, extracted DNA for next-generation sequencing, and assembled and validated the genome. ZEC performed data analysis and drafted the manuscript. HMC, QD and BW reviewed the manuscript critically for important intellectual content. All authors have read and agreed on the contents of the manuscript.

Funding This work was supported by the Chinese Academy of Medical Sciences, Innovation Funds for Medical Sciences (CIFMS) [2021-I2M-1-071 and 2021-I2M-1-022], National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705], National Science Foundation Funds [81872966], and Qinghai Provincial Key Laboratory of Phytochemistry of Qinghai Tibet Plateau [2020-ZJ-Y20]. The funders were not involved in the study design, data collection, analysis, decision to publish, or manuscript preparation.

Data availability The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (https://www.ncbi.nlm.nih.gov/) under the accession no. MW430049. The associated BioProject, BioSample and SRA numbers are PRJNA704532, SAMN18043523, and SRR13776395, respectively.

Declarations

Conflict of interest All the authors declare no conflicts of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

1. Liu C, Huang L-F (2020) Chloroplast genomic maps of Chinese medicinal plants [M], vol 1. Science Press, Beijing, China, pp 3–4
2. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:251364. https://doi.org/10.1155/2012/251364
3. Lohmann LG (2006) Untangling the phylogeny of neotropical lianas (Bignonieae, Bignoniaceae). Am J Bot 93(2):304–318. https://doi.org/10.3732/ajb.93.2.304
4. Flora of China (2021) Bignoniaceae. IBCAS Publishing iPlant. http://www.iplant.cn/info/Bignoniaceae?t=z. Accessed 07 Jan 2021
5. Flora of China (2021) *Campsis grandiflora*. IBCAS Publishing iPlant. http://www.iplant.cn/info/Campsis%20grandiflora?t=foc. Accessed 07 Jan 2021
6. Cui XY, Kim JH, Zhao X, Chen BQ, Lee BC, Pyo HB, Yun YP, Zhang YH (2006) Antioxidative and acute anti-inflammatory effects of *Campsis grandiflora* flower. J Ethnopharmacol 103(2):223–228. https://doi.org/10.1016/j.jep.2005.08.007
7. Commission CP (2015) The Pharmacopoeia of the People's Republic of China, 2015 edition part I. China Medical Science Press, Beijing
8. Xiao P-g, Zhao Z-z (2018) Encyclopedia of medicinal plants. World Book Inc, Beijing, China, pp 11–14
9. Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, McVeigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I (2020) NCBI taxonomy: a comprehensive update on curation, resources and tools. Database. https://doi.org/10.1093/database/baaa062
10. Steemers FJ, Gunderson KL (2005) Illumina, Inc. Pharmacogenomics 6:777–782. https://doi.org/10.2217/14622416.6.7.777
11. Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45(4):e18. https://doi.org/10.1093/nar/gkw955
12. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C (2019) CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res 47(W1):W65–W73. https://doi.org/10.1093/nar/gkz345
13. Hasegawa M, Kishino H, Saitou N (1991) On the maximum likelihood method in molecular phylogenetics. J Mol Evol 32(5):443–445. https://doi.org/10.1007/BF02101285
14. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32(1):268–274. https://doi.org/10.1093/molbev/msu300
15. Fonseca LHM, Lohmann LG (2017) Plastome rearrangements in the "*Adenocalymma-Neojobertia*" clade (Bignonieae, Bignoniaceae) and Its phylogenetic implications. Front Plant Sci 8:1875. https://doi.org/10.3389/fpls.2017.01875
16. Fonseca LHM, Lohmann LG (2018) Combining high-throughput sequencing and targeted loci data to infer the phylogeny of the "*Adenocalymma-Neojobertia*" clade (Bignonieae, Bignoniaceae). Mol Phylogenet Evol 123:1–15. https://doi.org/10.1016/j.ympev.2018.01.023
17. Thode VA, Lohmann LG (2019) Comparative chloroplast genomics at low taxonomic levels: a case study using *Amphilophium* (Bignonieae, Bignoniaceae). Front Plant Sci 10:796. https://doi.org/10.3389/fpls.2019.00796
18. Firetti F, Zuntini AR, Gaiarsa JW, Oliveira RS, Lohmann LG, Van Sluys MA (2017) Complete chloroplast genome sequences contribute to plant species delimitation: a case study of the *Anemopaegma* species complex. Am J Bot 104(10):1493–1509. https://doi.org/10.3732/ajb.1700302
19. Nazareno AG, Carlsen M, Lohmann LG (2015) Complete chloroplast genome of *Tanaecium tetragonolobum*: the first Bignoniaceae plastome. PLoS ONE 10(6):e0129930. https://doi.org/10.1371/journal.pone.0129930
20. Fonseca LH, Cabral SM, Agra Mde F, Lohmann LG (2015) Taxonomic updates in *Dolichandra* Cham. (Bignonieae, Bignoniaceae). PhytoKeys 46:35–43. https://doi.org/10.3897/phytokeys.46.8421
21. Jiang Y, Wang J, Qian J, Xu L, Duan B (2020) The complete chloroplast genome sequence of *Oroxylum indicum* (L.) Kurz

(Bignoniaceae) and its phylogenetic analysis. Mitochondrial DNA B 5(2):1429–1430. https://doi.org/10.1080/23802359.2020.1736961
22. Ma Q-g, Zhang J-g, Zhang J-p (2020) The complete chloroplast genome of *Catalpa ovata* G. Don.(Bignoniaceae). Mitochondrial DNA B 5(2):1800–1801. https://doi.org/10.1080/23802359.2020.1750979
23. Yang J, Wang S, Huang Z, Guo P (2020) The complete chloroplast genome sequence of *Catalpa bungei* (Bignoniaceae): a high-quality timber species from China. Mitochondrial DNA B 5(4):3854–3855. https://doi.org/10.1080/23802359.2020.1841581
24. Wu X, Li H, Chen S (2021) Characterization of the chloroplast genome and its inference on the phylogenetic position of *Incarvillea sinensis* Lam.(Bignoniaceae). Mitochondrial DNA B 6(1):263–264. https://doi.org/10.1080/23802359.2020.1860722
25. Ma G-T, Yang J-G, Zhang Y-F, Guan T-X (2019) Characterization of the complete chloroplast genome of *Incarvillea arguta* (Bignoniaceae). Mitochondrial DNA B 4(1):1603–1604. https://doi.org/10.1080/23802359.2019.1601529
26. Wu X, Peng C, Li Z, Chen S (2019) The complete plastome genome of *Incarvillea compacta* (Bignoniaceae), an alpine herb endemic to China. Mitochondrial DNA B 4(2):3786–3787. https://doi.org/10.1080/23802359.2019.1681916
27. Wang Y, Yuan X, Li Y, Zhang J (2019) The complete chloroplast genome sequence of *Spathodea campanulata*. Mitochondrial DNA B 4(2):3469–3470. https://doi.org/10.1080/23802359.2019.1674710
28. Yi D-K, Kim K-J (2016) Two complete chloroplast genome sequences of genus *Paulownia* (Paulowniaceae): *Paulownia coreana* and *P. tomentosa*. Mitochondrial DNA B 1(1):627–629. https://doi.org/10.1080/23802359.2016.1214546
29. Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res 6(5):283–290. https://doi.org/10.1093/dnares/6.5.283
30. Zhang D, Gao F, Jakovlic I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour 20(1):348–355. https://doi.org/10.1111/1755-0998.13096
31. Letunic I, Bork P (2007) Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23(1):127–128. https://doi.org/10.1093/nar/gkab301
32. Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISAweb: a web server for microsatellite prediction. Bioinformatics 33(16):2583–2585. https://doi.org/10.1093/bioinformatics/btx198
33. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27(2):573–580. https://doi.org/10.1093/nar/27.2.573
34. Gojman B, DeHon AVMATCH (2009) : Using logical variation to counteract physical variation in bottom-up, nanoscale systems. In: 2009 International Conference on Field-Programmable Technology, IEEE, pp 78-87. https://doi.org/10.1109/FPT.2009.5377684
35. Krumsiek J, Arnold R, Rattei T (2007) Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics 23(8):1026–1028. https://doi.org/10.1093/bioinformatics/btm039
36. Sullivan MJ, Petty NK, Beatson SA (2011) Easyfig: a genome comparison visualizer. Bioinformatics 27(7):1009–1010. https://doi.org/10.1093/bioinformatics/btr039
37. Amiryousefi A, Hyvönen J, Poczai P (2018) IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34(17):3030–3031. https://doi.org/10.1093/bioinformatics/bty220
38. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 32(5):1342–1353. https://doi.org/10.1093/molbev/msv022
39. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J (2006) KaKs_calculator: calculating Ka and Ks through model selection and model averaging. Genomics Proteomics Bioinform 4(4):259–263. https://doi.org/10.1016/S1672-0229(07)60007-2

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.