Phyton-International Journal of Experimental Botany
DOI:10.32604/phyton.2020.011211
Article

Comparative Analyses and Phylogenetic Relationships between *Cryptomeria fortunei* and Related Species Based on Complete Chloroplast Genomes

Yingting Zhang[1,2,3], Junjie Yang[1,2,3], Zhenhao Guo[1,2,3], Jiaxing Mo[1,2,3], Jiebing Cui[1,2,3], Hailiang Hu[1,2,3] and Jin Xu[1,2,3,*]

[1]Key Laboratory of Forest Genetics & Biotechnology of Ministry of Education, Nanjing Forestry University, Nanjing, 210037, China
[2]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, 210037, China
[3]College of Forestry, Nanjing Forestry University, Nanjing, 210037, China
*Corresponding Author: Jin Xu. Email: xjinhsh@njfu.edu.cn
Received: 27 April 2020; Accepted: 22 June 2020

Abstract: *Cryptomeria fortunei* (Chinese cedar) is a highly adaptable woody species and one of the main forest plantation trees in subtropical high-altitude areas in China. However, there are few studies on its chloroplast (cp) genome. In this study, the complete cp genome of *C. fortunei* was sequenced and evaluated via comparative analyses with those of related species (formerly the Taxodiaceae) in Cupressaceae. The *C. fortunei* cp genome was 131,580 bp in length, and the GC content of the whole genome was 35.38%. It lost one relevant large inverted repeat and contained 114 unique genes, including 82 protein-coding genes, 28 tRNAs and 4 rRNAs. The relative synonymous codon usage (RSCU) of codons ending with A/U was more than twice that of codons ending with G/C. Thirty long repeat structures (LRSs) and 213 simple sequence repeat (SSR) loci were detected in the *C. fortunei* cp genome. Comparative analyses of 10 cp genomes revealed that substantial rearrangements occurred in the gene organization. Additionally, 6 cp hotspot regions (*trnS-GGA*, *ycf1*, *trnP-GGG*, *trnC-GCA*, *psbZ* and *accD*) were identified, and 4 genes (*petL*, *psbM*, *rpl22* and *psaM*) had likely underwent positive selection. Phylogenetic analysis showed that Cupressaceae, Taxaceae and Cephalotaxaceae clustered to form a clade and that *C. fortunei* was most closely related to *C. japonica* (Japanese cedar), *C. japonica* cv. Wogon Hort and *Taxodium distichum* (baldcypress). These results provide references for future studies of population genetics, phylogenetic status and molecular markers among Cupressaceae species and for the cultivation of improved varieties.

Keywords: *Cryptomeria fortunei*; chloroplast genome; comparative analysis; phylogenetic analysis

1 Introduction

In terms of genera, as the most diverse conifer family, the modern Cupressaceae Bartling contains 135 extant species in 30 genera, including several genera formerly treated as Taxodiaceae Warming, such as *Glyptostrobus* Endl., *Sequoia* Endl., *Cryptomeria* D. Don, *Taxodium* Rich., *Metasequoia* Miki ex Hu et

Cheng, *Glyptostrobus* Endl. and *Taiwania* Hayata [1,2]. *Cryptomeria fortunei* Hooibrenk ex Otto et Dietris (Chinese cedar) is an evergreen coniferous species of the phylum Gymnospermae, class Coniferopsida, order Pinales and family Cupressaceae. It has become one of the main forest plantation species in subtropical high-altitude areas in China because of its rapid growth and high adaptability [3], and it has immense economic and ecological value. To date, studies on *C. fortunei* have mainly focused on analyses of chemical compositions [4] and forest cultivation [5], while its chloroplast (cp) genome has not yet been reported.

Biological characteristics are determined by heredity, while the origin of heredity is derived from complete genome sequences, including the nuclear genome and associated organelle genome (mitochondria/cp) sequences. The cp is one of the most essential organelles in green plants and algae. It carries out photosynthesis and several other biochemical pathways, including the biosynthesis of fatty acids and amino acids [6]. The complete cp genome of most plants is 120–180 kb in length and includes 100–120 genes [7]. In most angiosperms, cp genomes are maternally inherited. They usually contain a circular double-stranded DNA molecule and have a highly conserved organization consisting of two copies of a large inverted repeat (IR) region that separates the large and small single-copy (LSC and SSC) regions [8]. The cp genomes of gymnosperms have distinctive features, including paternal inheritance [9] and relatively high levels of intraspecific variations. IRs play an important role in stabilizing cp genomes [10]. Extensive losses of IR copies have often been observed in gymnosperms such as Pinaceae Lindl., Cupressaceae and the cupressophytes [11–14], which has led to some variations in genome structures and gene contents.

Since the tobacco cp genome was first reported in 1986 [15], the number of cp genome sequences has increased rapidly with the development of high-throughput sequencing technologies. To date, more than 1000 complete cp genomes have been sequenced, and the results are available in public sequence repositories [16]. Generally, elucidation of the molecular events in trees is difficult because of slow growth, long generation times and very large genomes [17]. However, cp genomes are relatively convenient to obtain, and their detection is relatively easy [18]; furthermore, cp genomes are widely used to study plant systematics, genetic improvements and phylogenomics for resolving complex evolutionary relationships [19–21]. Interpreting the complete sequences and conducting comparative analyses are effective and reliable means of accurately studying phylogeny, as shown in studies on Magnoliaceae Juss. [22], camellias [21], legumes [6] and pine trees [23,24]. Hao et al. [12] analysed cp genome gene content, organization, structure, IR loss and extensive cp genome rearrangements in *Glyptostrobus pensilis* (Staunt.) Koch (Chinese water pine) and other cupressophytes and found that they were similar. Zheng et al. [17] revisited the published complete *Cunninghamia lanceolata* (Lamb.) Hook (Chinese fir) and four other coniferous species cp genome sequences in Taxodiaceae and investigated the phylogenetic position of *C. fortunei* among conifers by examining gene functions, selection forces, substitution rates, and the full cp genome sequence. Nevertheless, relatively complete comparative analyses of the cp genomes of *C. fortunei* and related species (formerly Taxodiaceae) have rarely been reported and remain to be explored.

In this study, we reported the complete sequence of the *C. fortunei* cp genome and then presented the results of comparative genomic analyses of cp genome sequences from related species in Cupressaceae, including analyses of the gene contents, codon usage, repeated sequences, nucleotide diversity (pi), Ka/Ks ratios and rearrangements. A phylogenetic analysis was performed on the basis of the whole cp genome of 37 Gymnosperm species to improve the understanding of the cp genome evolution of Cupressaceae plants. These results are expected to reveal structural features, enable the exploration of hypervariable regions for the identification of potential DNA barcodes, and provide a theoretical basis for the determination of phylogenetic statuses and future scientific studies.

2 Materials and Methods

2.1 Plant Material, DNA Extraction, Sequencing and Genome Assembly

Needles of *C. fortunei* were collected at Nanjing Forestry University (118°50'E, 32°05'N) and quickly stored at –80°C until analysis. Approximately 2 g of sample was used to extract DNA via a modified cetyltrimethylammonium bromide (CTAB) method [25]. Agarose gel electrophoresis and spectrophotometry were used to assess the DNA quality and quantity, respectively. Short-insert paired-end libraries were constructed from pure DNA. Approximately 5 μg of total DNA was pooled and run in a single lane of an Illumina HiSeq 4000 platform (San Diego, CA, USA) with a read length of 150 bp. Our assembly process was divided into the following 5 steps. First, SPAdes version 3.10.1 was used to assemble cp DNA sequences to obtain seed sequences [26]. Second, short seed kmers were iteratively extended into contigs. Third, SSPACE version 2.0 was used to connect the obtained contigs to scaffold sequences [27]. Fourth, GapFiller version 2.1.1 was used to complement the gaps of the obtained scaffold sequences until a complete pseudo-genome sequence was obtained [28]. Fifth, the sequence was aligned to the pseudo-genome and then corrected by genomic recombination. Finally, coordinates were rearranged according to the structure of the cp genome to obtain a complete circular cp genome sequence.

2.2 Genome Annotation and Analysis of cp DNA Sequences

BLAST version 2.2.25 was used to compare the coding sequences (CDSs) of the cp genomes, and the genes of the cp genomes were annotated using CpGAVAS [29] after manual adjustment. HMMER version 3.1b2 [30] and ARAGORN version 1.2.38 [31] were used to annotate the ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) in the cp genomes. A circular gene map of *C. fortunei* was drawn using the OGDRAW program [32]. An analysis of variation in relative synonymous codon usage (RSCU) and GC contents of complete cp genomes was conducted [33]. MISA version 1.0 [34] was used to visualize the simple sequence repeats (SSRs). The minimum numbers for the SSR motifs were 5, 5, 3, 3, 3 and 3 for mono-, di-, tri-, tetra, penta- and hexanucleotide repeats, respectively. REPuter [35] was used to visualize long repeat structures (LRSs). The constraints were set as follows: ≥90% sequence identity based on a Hamming distance of 3 and n ≥30 bp. MAFFT version 7.310 was used to extract and align separate CDSs of the same protein-coding gene [36]. The proportion of mutation events (%) = [(NS + ID)/L] × 100, where NS = the number of nucleotide substitutions, ID = the number of indexes, and L = the sequence length of the arrangement. KaKs Calculator version 2.0 [37] with the MLWL model [38] and vcftools [39] were used to calculate the Ka/Ks ratios and pi values of the genes, respectively.

2.3 Genome Comparison

Rearrangement analyses of ten cp genomes were performed using Mauve version 2.3.1 alignment [40]. CGVIEW [41] was used for comparative analyses of the cp genome structure of *C. fortunei* and related species, i.e., *Sequoia sempervirens* (Lamb.) Endl (redwood), *Taxodium distichum* (L.) Rich (baldcypress), *C. japonica* (L. f.) D. Don (Japanese cedar), *C. japonica* cv. Wogon Hort, *C. lanceolata*, *Metasequoia glyptostroboides* Hu et Cheng (dawn redwood), *G. pensilis*, *Taiwania* (*Taiwania cryptomerioides* Hayata and *Taiwania flousiana* Gaussen).

2.4 Phylogenetic Analysis

A phylogenetic tree was constructed by maximum-likelihood (ML) analysis using the 37 cp genome sequences of gymnosperm species from the National Center for Biotechnology Information (NCBI) Organelle Genome and Nucleotide Resources database. The sequences were aligned by using MAFFT [36], and then visualization and manual adjustment of multiple sequence alignments were conducted in BioEdit [42]. The phylogenetic tree was constructed by RAxML (version 8.2.10, https://cme.h-its.org/

exelixis/software.html), which uses a rapid hill-climbing algorithm and the general time reversible (GTR) model for the ML analysis. The local bootstrap probability of each branch was calculated by 100 replications.

3 Results

3.1 The cp Genome Size and Features of C. fortunei

The completed cp genome sequence of *C. fortunei* was submitted to GenBank (accession number MN509811). The complete size of the *C. fortunei* cp genome was 131,580 bp (Fig. 1), slightly larger than the cp genomes of *T. cryptomerioides* (131,427 bp) and *T. flousiana* (131,413 bp) but smaller than the cp genomes of other related species (Tab. 1), which are similar to the lengths of the cp genomes of most higher plants (between 120 and 160 kb) [43]. Large IR regions were not detected in the *C. fortunei* cp genome; therefore, we were unable to define the LSC and SSC regions in this genome. The GC content is an important indicator of species affinity [44]. The GC content of the whole *C. fortunei* cp genome was 35.38%, and the GC content of the cp genomes of related plants ranged from 34.72% to 35.38% (Tab. 1), which were similar to those of *Pinus taeda* L. (Loblolly pine, 38.50%) [45], *Agathis dammara* (Lamb.) Rich. et A. Rich (Queensland kauripine, 36.54%), *Nageia nagi* (Thunb.) O. Kuntze (Podocarpus nagi, 37.26%) and *Calocedrus formosana* (Florin) Florin (Taiwan incense cedar, 34.83%) [46].

A total of 120 genes were identified in the *C. fortunei* cp genome, slightly more than the number of genes identified in its related species (113–119 genes) (Tab. 1). Plant cp genomes may have 63–209 genes but primarily have 110–130 genes [47]. These genes are highly conserved among land plant plastomes [48]. A total of 114 genes were single copy or multi-copy, of which 4 were rRNA genes (3.51%), 28 were individual tRNA genes (24.56%), and 82 were protein-coding genes (71.93%): 4 genes encoding DNA-dependent RNA polymerases, 21 genes encoding small and large ribosomal subunits, 45 genes encoding photosynthesis-related proteins, and 12 genes encoding other proteins, including 3 hypothetical genes (*ycf3, ycf2* and *ycf1*) (Tabs. 1 and 2). *TrnS-UUA* (2401 bp) and *ycf2* (33 bp) had the largest and smallest introns, respectively (Appendix A). Additionally, there were 15 intron-containing genes (9 protein-coding genes and 6 tRNA genes), of which 14 genes contained one intron and only the *ycf3* gene contained two introns. Introns and exons play an important role in the transcriptional regulation of genes [22,49]. Studies have shown that *ycf3* is required for the stable accumulation of photosystem I complexes [50]. Therefore, we speculate that the gain of two introns in *ycf3* of *C. fortunei* may be helpful in studying the mechanism of photosynthesis evolution. In general, the gene content and organization of cp genomes are relatively conserved.

In the *C. fortunei* cp genome, the protein CDSs were composed of 74,607 bp, accounting for 56.71% of the total cp genome; the gene proportions for tRNA and rRNA were 1.91% and 7.66%, respectively. A total of 43.35% of the non-coding regions were composed of introns and intergenic spacers. Based on the CDSs of *C. fortunei*, the RSCU values ranged from 0.34 to 2.98, and the RSCU values of 31 codons were greater than 1 (Appendix B). Specifically, UUA, GCU, AGA, UCU, ACU, CCU, GGA, UAA, GAU, UAU, CAA, GCU, GAA, AAU, AAA, GUA, AUU, UGU, CGU, GGU, GUU, UUU, CGA, AGU, UUG, CUU, GCA, ACA, UCA, CCA and AUA were the preferred codons. The RSCU values of AUG and UGG were equal to 1 (Appendix B). Usage of the start, tryptophan UGG and methionine AUG codons had no bias [51]. The RSCU values of the 64 codons of *C. fortunei* cp genomes exhibited little variance; therefore, we calculated the RSCU values of codons ending with A/U and with C/G. The total RSCU values of codons ending with A/U and G/C in *C. fortunei* were 44.89 and 18.95, respectively (Appendix B). Codon usage and mutational bias play an important role in shaping cp genome evolution [52]. These results indicate that the most preferred synonymous codons ended with A/U [14]. Specifically, codon usage was biased towards a high representation of A/U at the third codon position [53,54]. These results have significance for guiding the identification of unknown genes in the genome.

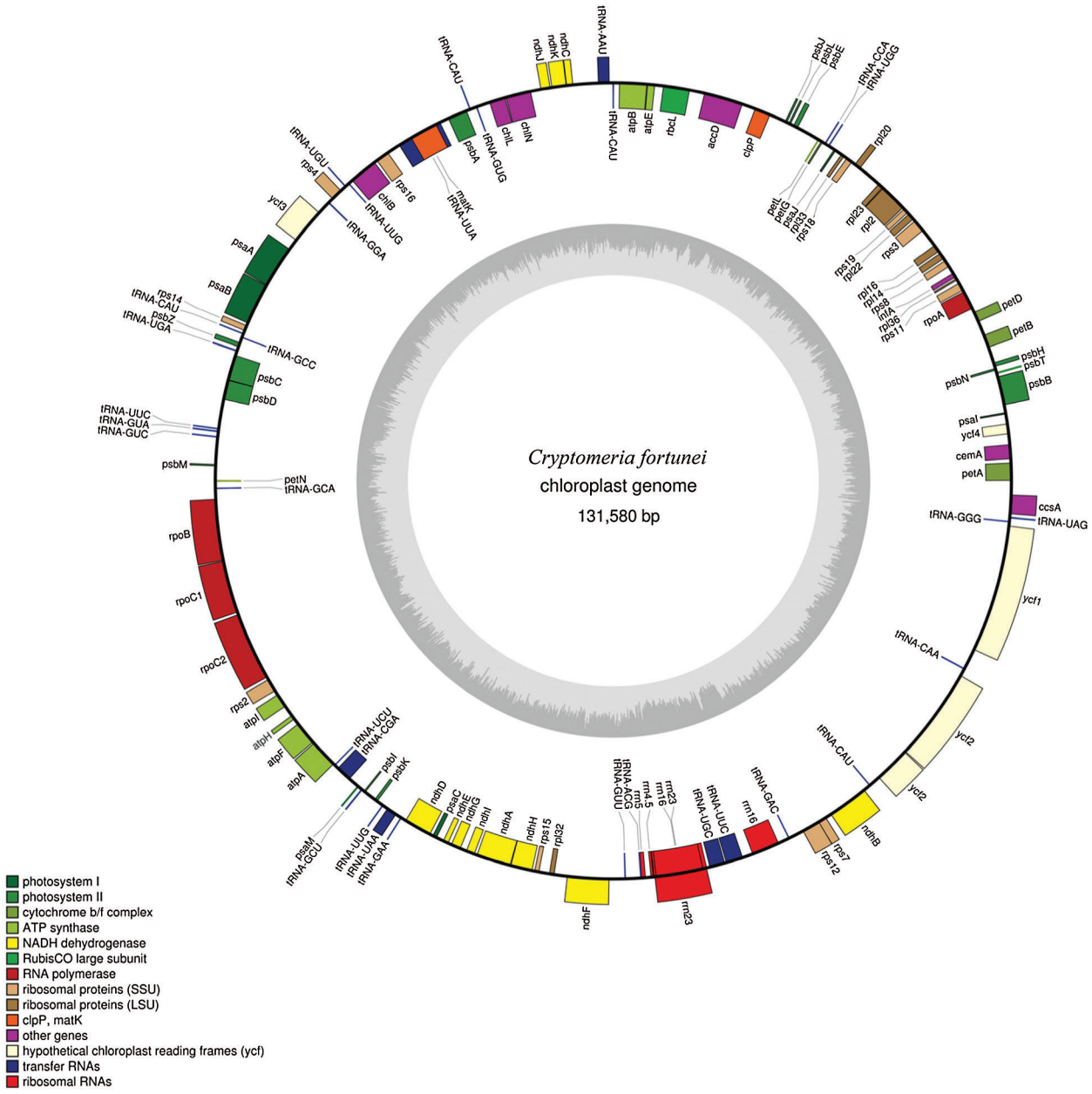

Figure 1: Chloroplast gene map of *C. fortunei*. Genes outside the circle are transcribed clockwise, whereas genes inside the circle are transcribed counterclockwise. The light grey region indicates the AT content, whereas the dark grey region in the inner circle indicates the GC content

3.2 Analysis of Long Repeat Structures (LRSs)

LRSs are repeat sequences that are 30 bp or longer, and they are classified into 4 types: forward, reverse, complementary and palindromic repeats [29]. In the *C. fortunei* cp genome, 30 LRSs were detected, including 18 forward and 12 palindromic repeats, of which 5 LRSs were located on the genes *rps18*, *chlB*, *trnA-CAU*, *ycf1* and *ycf2* (Appendix C). Among 10 cp genomes, 309 LRSs contained 221 forward and 88 palindromic repeats, and *M. glyptostroboides* and *S. sempervirens* had the fewest (6 forward and 5 palindromic repeats) and most (44) LRSs, respectively (Fig. 2; Appendix D). The length of the LRSs in

Table 1: Features of the chloroplast genomes of *C. fortunei* and related species

Species	Accession number	Length (bp)	GC (%)	Number of genes			
				tRNA	rRNA	mRNA	Total
C. fortunei	MN509811	131,580	35.38	32	6	82	120
C. japonica	NC_010548.1	131,810	35.38	32	4	82	118
C. japonica cv.Wogon Hort	AP010966.1	131,804	35.38	32	4	81	117
C. lanceolata	NC_021437.1	135,334	35.00	35	0	81	116
G. pensilis	NC_031354.1	132,239	35.31	32	4	83	119
M. glyptostroboides	NC_027423.1	131,887	35.25	31	4	82	117
S. sempervirens	NC_030372.1	133,929	35.37	34	4	75	113
T. cryptomerioides	MG963174.1	131,427	34.72	34	0	83	117
T. flousiana	NC_021441.1	131,413	34.72	34	0	83	117
T. distichum	NC_034941.1	131,954	35.26	32	4	82	118

Table 2: Genes in the sequenced *C. fortunei* chloroplast genome

Function	Gene name
RNAs, transfer	*trnA-ACG, trnA-UCU, trnA-GUC, trnA-GUU, trnA-UGC*, trnC-GCA, trnG-GCC, trnG-UUG* × 2, *trnG-UUC, trnG-UUC*, trnH-GUG, trnI-AAU*, trnL-CAA, trnL-UAA*, trnL-UAG, trnM-CAU* × 4, *trnT-CCA, trnT-GUA, trnP-GAA, trnP-GGG, trnP-UGG, trnS-CGA*, trnS-GCU, trnS-GGA, trnS-UGA, trnS-UUA*, trnT-UGU, trnV-GAC*
RNAs, ribosomal	*rrn4.5, rrn5, rrn16* × 2, *rrn23* × 2
Transcription and splicing	*rpoA, rpoB, rpoC1*, rpoC2*
Translation, ribosomal proteins	
Small subunit	*rps2, rps3, rps4, rps7, rps8, rps11, rps12*, rps14, rps15, rps16*, rps18, rps19*
Large subunit	*rpl2*, rpl14, rpl16, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36*
Photosynthesis	
ATP synthase	*atpA, atpB, atpE, atpF*, atpH, atpI*
Photosystem I	*psaA, psaB, psaC, psaI, psaJ, psaM, ycf4*
Photosystem II	*psbA, psbB, psbC, psbD, psbE, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ*
Calvin cycle	*rbcL*
Cytochrome complex	*petA, petB, petD, petG, petL, petN*
NADH dehydrogenase	*ndhA*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndh, ndhI, ndhJ, ndhK*
Other genes	*matK, infA, cemA, accD, ccsA, chlN, chlL, chlB, clpP, ycf3**, ycf2*, ycf1*

Note: *Genes containing one intron; **genes containing two introns

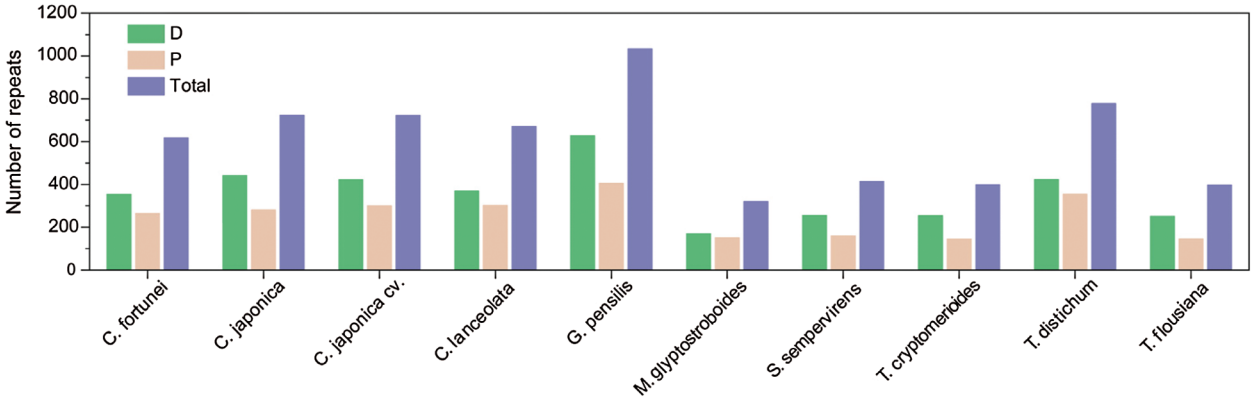

Figure 2: Analysis of the type of long repeat structures (LRSs) of the chloroplast genome in *C. fortunei* and related species. D, forward repeats; P, palindromic repeats (including inverted repeats and complementary sequences). *C. japonica* cv., *C. japonica* cv. Wogon Hort

the *C. fortunei* genome ranged from 30 to 275 bp but was primarily between 30 and 47 bp, which was also similar to the values in the cp genomes of its related species (Appendix D).

3.3 Analysis of Simple Sequence Repeats (SSRs)

SSRs are sequences containing repeated motifs 1–6 bp in length. A total of 213 SSRs were found in the *C. fortunei* cp genome, including 122 (57.28%) mono-, 23 (10.80%) di-, 57 (26.76%) tri- and 11 (5.16%) tetranucleotides repeats (Fig. 3A), which was substantially larger than the number of SSRs found in Magnoliaceae (average each species 50 SSRs) [22], *Quercus acutissima* Carruth (sawtooth oak, 65 SSRs) [33], *Camellia* (each 50–55 SSRs) [21] and *P. taeda* (151 SSRs) [45] and almost the same number as that in *Anthriscus cerefolium* L. Hoffm. (217 SSRs) [53]. These were also similar to the proportions in the 10 cp genomes, in which the numbers of mono-, di-, tri-, tetra-, penta- and hexanucleotides were 1169 (56.75%), 209 (10.15%), 582 (28.25%), 90 (4.37%), 8 (0.39%) and 2 (0.10%), respectively (Fig. 3A). In addition, non-coding regions (intergenic regions) and introns in coding regions of *C. fortunei* contained 146 SSRs, which was similar to the distribution of SSRs in the 10 cp genomes (1255 SSRs) (Fig. 3B), and only a small number of SSRs (67 SSRs) were distributed in exon regions (Fig. 3B). The genes with 3 or more SSRs were *rpoB, rpoC1, rpoC2, rps2, ndhD, ndhA, accD, matK, ycf2* and *ycf1* (Appendix E). In *C. fortunei*, a total of 118 SSRs were made up of only single A/T bases, and 21 SSRs contained AT/TA, accounting for 65.26% of the total SSRs (Appendix F and G). Therefore, A and T were abundant in SSRs and accounted for 84.07%, which was similar to the composition of SSRs in its related species (80.47–84.77%) (Appendix F and G).

3.4 Analysis of Nucleotide Variability (pi) in the Ten Cp Genomes

For all genes, the pi values ranged from 0 to 0.25, and the mean value was 0.030. Among them, the mean values of tRNAs, encoding genes and rRNAs were 0.0241, 0.034 and 0.003, respectively (Fig. 4). Six genes (*trnS-GGA, ycf1, trnP-GGG, trnC-GCA, psbZ* and *accD*) had pi values > 0.1, and *trnS-GGA* had the highest pi (0.25); *ycf2, rpl22, clpP* and *rpl32* followed with the next highest pi values. Nine genes (*trnA-ACG, rrn5, trnM-CAU, trnL-UAG, trnG-UUC, trnA-GUC, trnS-CGA, trnL-UAA, trnP-GAA* and *trnS-UUA*) had pi values of 0 (Fig. 4).

3.5 Ka/Ks Analysis

To test the differences in the evolutionary rates of cp genomes in *C. fortunei* and related species, we examined the sequence divergences of non-synonymous (Ka) and synonymous (Ks) substitution rates.

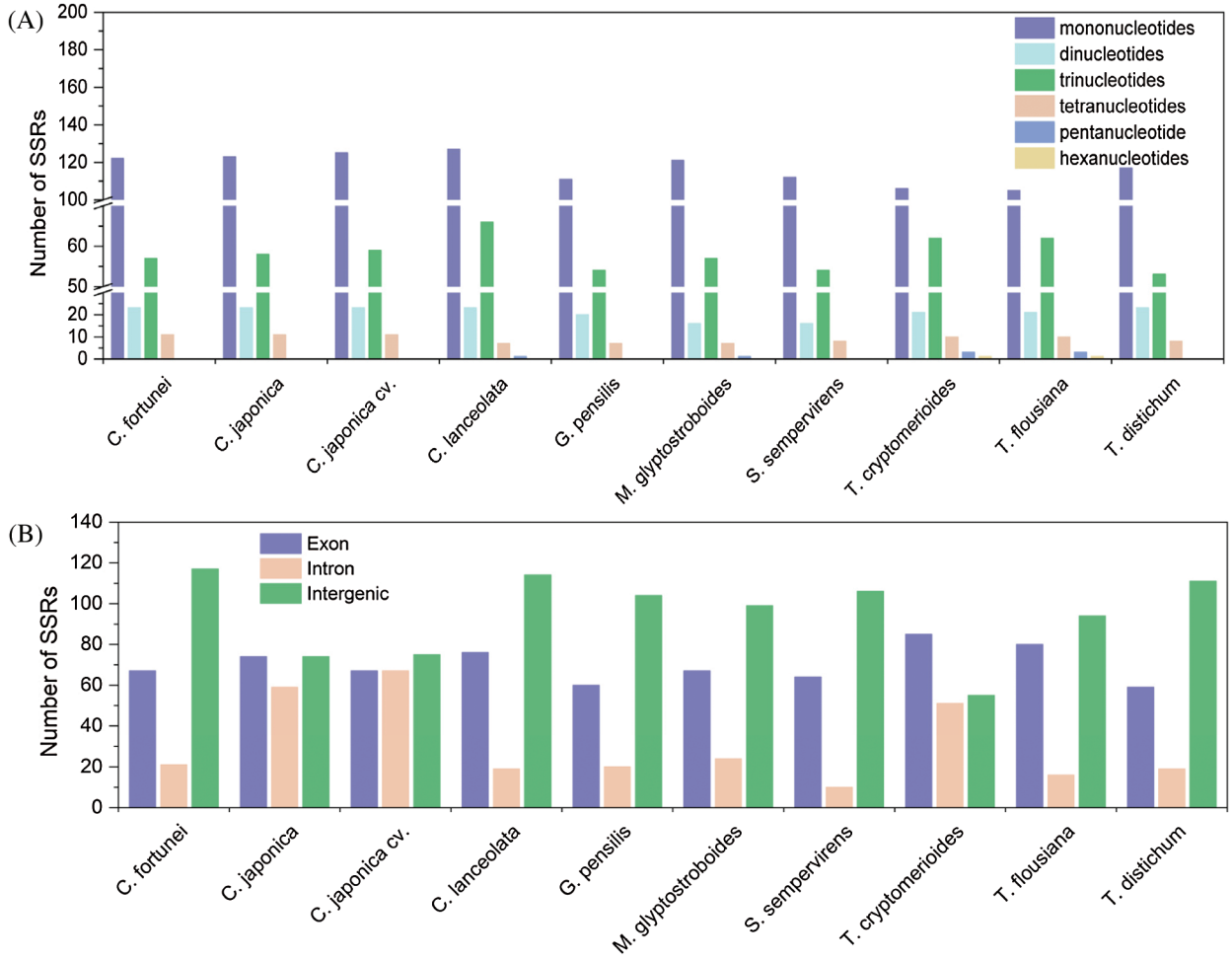

Figure 3: Comparison of simple sequence repeat (SSR) distributions of chloroplast genomes between *C. fortunei* and related species. (A) Number of different SSR types detected in the ten chloroplast genomes; (B) Frequency of SSRs in the intergenic regions, exons and introns, *C. japonica* cv., *C. japonica* cv. Wogon Hort

The average Ka/Ks ratio of 80 protein-coding genes was determined to be 0.41 by comparing the *C. japonica* genome and related genomes. There were only 3 genes with nonzero Ka/Ks ratios between *C. fortunei* and *C. japonica* or *C. japonica* cv. Wogon Hort, whereas other species had 51–69 genes. The Ka/Ks ratios of most genes were less than 1, of which 62 genes (77.50%) had values less than 0.5 (Appendix H). The genes with mean Ka/Ks ratios between 0 and 0.01 were *atpH*, *petG*, *petN*, *psaC*, *psbL* and *psbN*, and the genes with average Ka/Ks > 1 were *petL*, *psbM*, *rpl22* and *psaM* (Appendix H).

3.6 Cp Genome Rearrangements among C. fortunei and Related Species

The variation in the non-coding region was obviously higher than that in the coding region among these 10 cp genomes. The cp genome of *C. fortunei* was most similar to those of the same genus, such as *C. japonica* and its cultivated species, and was greatly different from those of *M. glyptostroboides, G. pensilis* and *S. sempervirens*. There were also some differences among the remaining species genomes. Overall, except for plants of the same genus, the cp genome sequences showed substantial rearrangements among Cupressaceae species (Fig. 5).

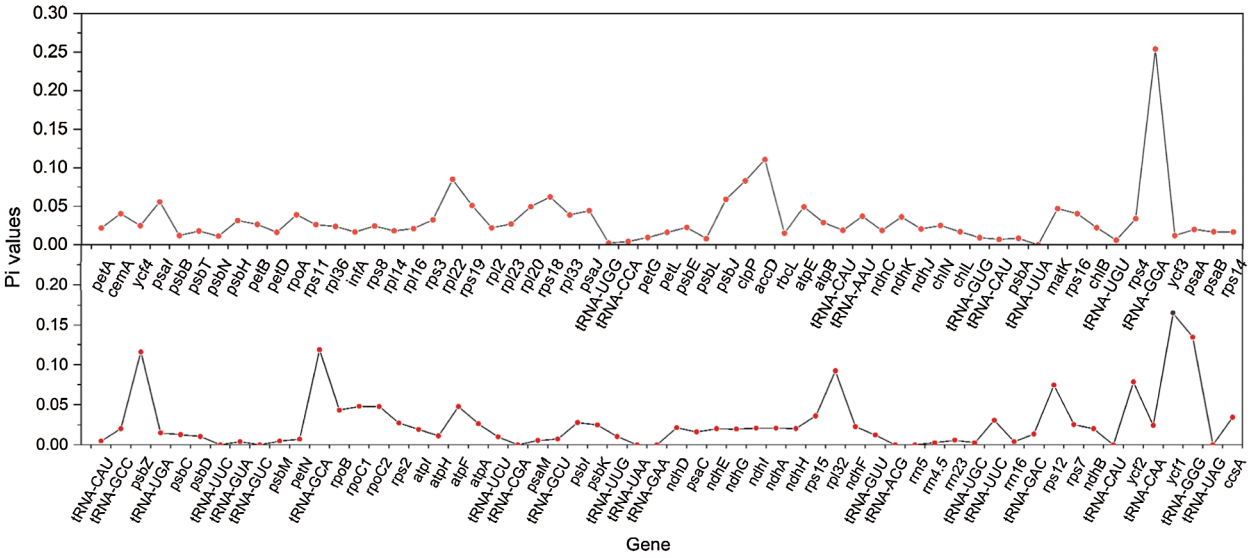

Figure 4: Analysis of the nucleotide variability (pi) values of the whole chloroplast genomes of *C. fortunei* and related species

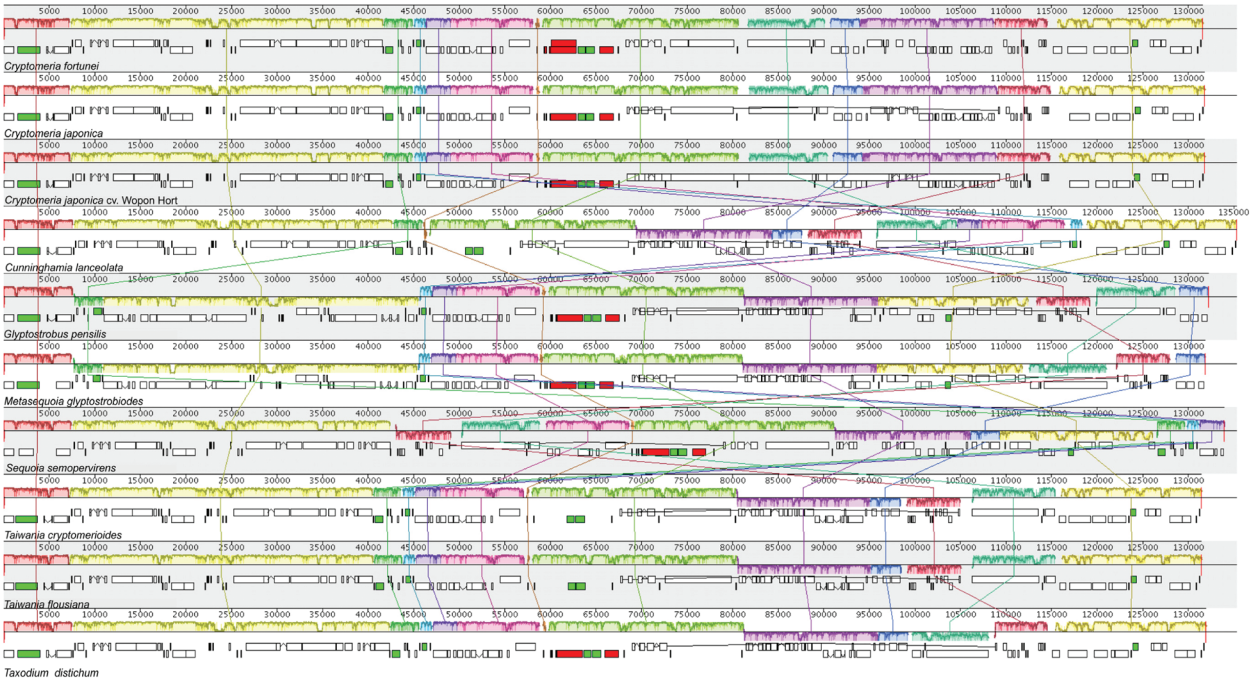

Figure 5: Mauve alignment of *C. fortunei* and related species' genomes. The *C. fortunei* genome is shown at top as the reference. Within each alignment, local collinear blocks (long squares) are represented by the same coloured blocks connected by lines, of which white, green and red represent CDSs, tRNAs and rRNAs, respectively. Short rectangles represent the genetic location of each genome

3.7 Phylogenetic Analysis

In general, support was high for almost all relationships inferred from the data for 37 cp genomes based on ML methods with very high overall bootstrap values (Fig. 6). *Ginkgo biloba* L. (maidenhair tree) was set

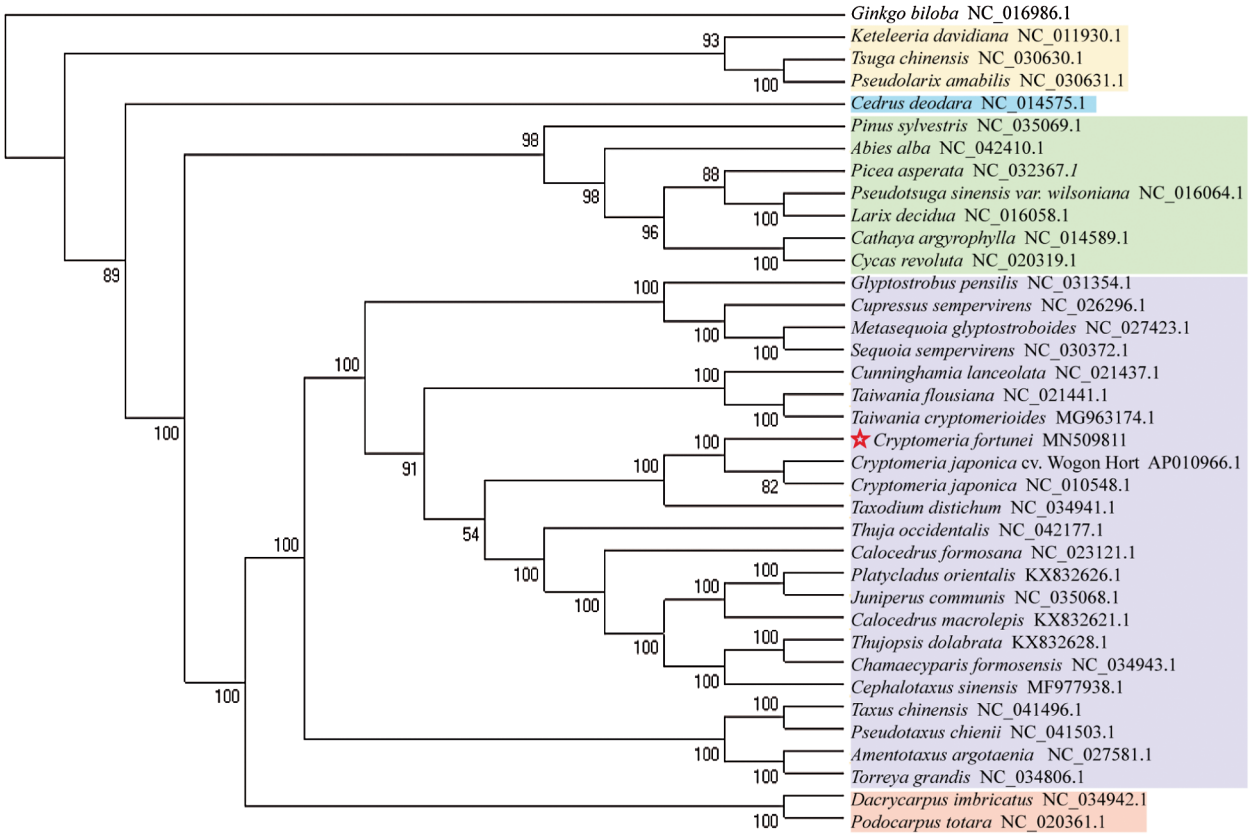

Figure 6: Phylogenetic tree reconstruction for 37 gymnosperm species. ML analysis based on all chloroplast genomes with the GTR model. Numbers above and under the nodes are bootstrap values. *Ginkgo biloba* was set as the outgroup

as an outgroup for the construction of our phylogenetic tree. The phylogenetic tree was divided into five clades. Pinaceae and *C. revoluta* dispersed into the first three clades. Podocarpaceae formed a clade alone. Cupressaceae, *Cephalotaxus sinensis* (Rehd. et Wils.) Li (China Plumyew, Cephalotaxaceae Neger) and Taxaceae S. F. Greyclustered to form a clade. In addition, *C. fortunei* was most closely related to *C. japonica*, *C. japonica* cv. Wogon Hort and *T. distichum*. *M. glyptostroboides* and *S. sempervirens* were also closely related to each other, as were *C. lanceolata* and *Taiwania* (Fig. 6).

4 Discussion

4.1 Lost IR Regions and Structural Differences in C. fortunei and Related Species in Cupressaceae

We found that in *C. fortunei*, the structure of the complete IRs was lost from the cp genome, a finding that was consistent with results of previous studies of its related species, such as *G. pensilis*, *M. glyptostroboides*, and *C. lanceolata* [12,13,17]. We also found that y*cf1*, *ycf2* and *ycf3* lost only some homologous sequences and formed pseudogenes because of IR loss, similar to results from studies on *ycf2* [55,56]. Large IRs are thought to be correctly copied and stabilize the cp genomes against rearrangements [57]. Specifically, species that have lost their IRs usually have undergone more rearrangements than those that have not [58]. For instance, 4 algal cp genomes showed a strong correlation between the number of repeats and the degree of rearrangements [59,60]. Thus, the most highly rearranged green algal cp genome has the greatest number of repeats in its lineage. We speculated that cp genomic rearrangement clearly occurred in different genera formerly of Taxodiaceae due to the

loss of large IRs [17,61]. Rearrangements in cp genomes may be useful as phylogenetic markers within genera or even within families, becoming a potential tool for understanding the evolution of plant species [62]. In addition, rearrangements of cp DNA fragments or genes have been considered evolutionarily rare and are often mapped on phylogenetic trees to support particular tree nodes.

4.2 Cp Genome Markers and Divergence Hotspots among C. fortunei and Related Species

LRSs may promote the rearrangement of the cp genomes and increase the genetic diversity of populations [53]. In this study, 309 pairs of LRSs were detected in ten cp genomes, including 30 LRSs in the *C. fortunei* cp genome that were primarily 30–47 bp in length, similar to the number and length of LRSs in Magnoliaceae [22]. The *C. fortunei* cp genome had few LRSs located on genes, which was similar to the situation in other angiosperms in which most repeats were located in intergenic regions or *ycf* (*ycf1* and *ycf2*) genes [63,64]. The existence of these repeats implies that the region is a potential hotspot for genomic rearrangement [65] and may serve as a genetic marker for population genetics and phylogenetic studies.

SSRs are efficient molecular markers with high abundance, high polymorphism, co-dominant inheritance, reproducibility and transferability; therefore, they have been widely used in species identification, population genetics, the construction of genetic linkage maps and evolutionary studies [66–72]. In *C. fortunei*, most SSRs were only A/T mononucleotides, similar to the SSRs in *Quillaja saponaria* Molina (soapbark), in which the most common SSRs are A/T mononucleotide repeats [73]. To an extent, AT content is related to stability [74]. Thus, SSRs are generally composed of AT repeats and are rich in AT in cp genomes. In *C. fortunei*, 67 SSRs (31.46%) were observed in the exon region (no SSRs were found in the tRNAs and rRNAs). These results are also consistent with findings in *Salvia miltiorrhiza* Bunge (red or Chinese sage), which has unevenly distributed SSRs within its cp genome and fewer SSRs in non-coding regions, as there are more highly variable regions in the cp genomes [75] than in exons [53]. In *C. fortunei*, we found that some genes contained two or more SSR motifs, such as *rpoB, rpoC1, rpoC2, ndhD, ndhA, matK, ycf2* and *ycf1*, which was similar to findings in *Ananas comosus* (Linn.) Merr. (pineapple) [76]. In recent years, genomic SSR markers have received increasing attention due to the detection of higher levels of polymorphism in SSRs than in EST-SSRs and increased variable introns or intergenic sequences [77]. In future studies, these cp SSR markers could be used to examine genetic structure, diversity and differentiation in *C. fortunei* and related species.

Pi values can reveal variations in nucleic acid sequences in different species. The highly variable regions reported can be used as potential molecular markers for population genetics, phylogenetic analysis and species identification [78,79]. In the 10 cp genomes, we found that the average value of 4 rRNAs was the lowest (0.003), which may indicate that these are the most highly conserved genes. We also found that some encoding regions differed among these species, e.g., *trnS-GGA, ycf1, trnP-GGG, trnC-GCA, psbZ* and *accD* (Pi > 0.1), which was consistent with results from *Meconopsis* [54]. These gene sequences can be used as hotspot regions and DNA markers for classification and revealing the genetic divergence of Cupressaceae.

Previous studies have demonstrated that the *ycf1* gene has high divergence in cp genomes and is recommended as a core DNA barcode for plants [80]. It has been increasingly widely applied in plant DNA barcode studies [81]. In our study, *ycf1* was in LRSs and SSRs and had high pi value. Therefore, we speculated that *ycf1* can be used for the cp DNA barcode of Cupressaceae. In addition, we newly identified SSRs and LRSs of *C. fortunei* and other high-pi genes of Cupressaceae that may also be used as potential molecular markers and suitable barcodes for plant classification in Cupressaceae.

4.3 Evolutionary Pressure and Phylogenetic Evolution of cp Genomes in Cupressaceae

We found that in 10 cp genomes, the Ka/Ks ratios of most genes were less than 0.5, and those of several genes were less than 0.01, which is consistent with the results for green plants [59], suggesting that most

DNA genes exhibited purifying selection (Ka/Ks < 1 or 0.5) [60] or strong purifying selection pressure (0–0.01) [61]. Among these genes, the Ka/Ks ratios of 4 genes (*petL*, *psbM*, *rpl22* and *psaM*) were greater than 1, indicating probable positive selection [59,60]. These genes may be in rearrangement regions [62], play an important role in the evolution of species and indicate specific proteins within a broader group of species. Ka/Ks < 1 (especially less than 0.5) indicates purifying selection [82], and average Ka/Ks ratios between 0 and 0.01 indicate very strong purifying selection pressure [83]. In this study, the Ka/Ks ratios of 62 genes were less than 0.5, and those of 6 genes (*atpH*, *petG*, *petN*, *psaC*, *psbL* and *psbN*) were less than 0.01, suggesting that cp DNA genes exhibited purifying selection and were highly conserved, which is consistent with green plants [84].

Cp genome sequences have been widely used for the reconstruction of phylogenetic relationships among plant lineages [22,85,86]. Surviving former Taxodiaceae plants belong to monotypic or oligotypic genera, and differences between genera are very obvious. To date, Taxodiaceae and Cupressaceae still lack an accepted classification system for phylogenetic relationships. In this study, Cupressaceae (including formerly Taxodiaceae), *Cephalotaxus sinensis* and Taxaceae clustered to form one clade, which was similar to results from previous studies indicating that Taxodiaceae, Taxaceae and Cephalotaxaceae should be included in Cupressaceae [87–90]. *C. fortunei* was most closely related to *C. japonica*, *C. japonica* cv. Wogon Hort and *T. distichum*. *M. glyptostroboides* and *S. sempervirens* were also closely related to each other. These results were consistent with previous studies that showed that Taxodiaceae (*Cryptomeria*, *Taxodium and Glyptostrobus*) and Sequoioideae (*Metasequoia* and *Sequoia*) form monophyletic groups [91]. The interplay between enhancer and coding sequence evolution created a potentially adaptive path for morphological evolution [92]. Interestingly, there were few differences in *Glyptostrobus* species. Additionally, most Chinese scholars still use *C. fortunei* [93–95], which is controversial. Tsumura et al. [96] investigated molecular phylogeny based on six polymerase chain reaction-amplified cp genes (*frxC*, *rbcL*, *psbA*, *psbD*, *trnK* and *16S*) and found *C. japonica* is close *to C. fortunei*. Tsumura et al. [97] investigated population samples for three *C. japonica* groups (including *C. fortunei*, namely *C. japonica* var. sinensis) using high-throughput SNP genotyping and sequencing of multiple genes and then estimating the level of pi, the genetic differentiation, and divergence time, which led them to find that climate change plays an important role in speciation processes of *C. japonica*. We also found that they had some small differences in the number of rRNAs, SSRs, LRSs, etc. It is possible that the gene content and organization of cp genomes are relatively conserved. The classification of *C. japonica* and *C. fortunei* maybe need more experiments to be further explored. And phylogenetic trees are based on a certain criterion; thus, whether phylogenetic trees can widely correctly represent gymnosperm species remains to be further verified.

5 Conclusions

In this study, we first reported the complete cp genome of *C. fortunei* and analysed the cp genome features, LRSs, SSRs, codon usage, pi, Ka/Ks ratios and phylogenetic relationships of *C. fortunei* and related species. In 10 cp genomes, substantial rearrangements occurred in gene organization because of IR loss, 6 cp hotspot regions were identified and 4 genes were related to probable positive selection. The cp genome was shown to be more conservative with similar characteristics to the genomes of related species. Understanding the phylogenetic relationships of Cupressaceae species (including formerly Taxodiaceae) has gradually improved through phylogenetic analysis. However, the location of *Glyptostrobus* requires more data to support our findings. These results will facilitate further studies on population genetics, breeding, molecular markers and phylogenies in Cupressaceae.

Funding Statement: This research was funded by the National Forestry and Grassland Administration of China, National Forestry Public Welfare Industry Research Project (Grant No. 201304104), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Li, L., Jin, J., Manchester, S. R. (2018). Cupressaceae fossil remains from the Paleocene of Carneyville, Wyoming. *Review of Palaeobotany and Palynology, 251,* 1–13. DOI 10.1016/j.revpalbo.2017.12.003.
2. Yu, Y. F. (1995). Origin, evolution and distribution of Taxodiaceae. *Journal of Systematics and Evolution, 33,* 362–339 (in Chinese).
3. Lu, Y., Zhang, J., Tong, Z. (2015). Exploit SSR in *Cryptromeria fortunei* and application in genetic diversity analysis. *Journal of Forestry Research, 26(4),* 963–969. DOI 10.1007/s11676-015-0093-1.
4. Xie, Y., Huang, Q., Yang, F., Lei, C. (2012). Chemical variation in essential oil of *Cryptomeria fortunei* from various areas of China. *Industrial Crops and Products, 36(1),* 308–312. DOI 10.1016/j.indcrop.2011.10.023.
5. Hao, Y. Q. (2006). Preview of spatial structure of *Cryptomeria fortunei* plantation after stand improvement. *Scientia Silvae Sinicae, 8,* 8–13.
6. Saski, C., Lee, S., Daniell, H., Wood, T. C., Tomkins, J. et al. (2005). Complete chloroplast genome sequence of *Glycine max* and comparative analyses with other legume genomes. *Plant Molecular Biology, 59(2),* 309–322. DOI 10.1007/s11103-005-8882-0.
7. Palmer, J. D. (1985). Comparative organization of chloroplast genomes. *Annual Review of Genetics, 19(1),* 325–354. DOI 10.1146/annurev.ge.19.120185.001545.
8. Palmer, J. D. (1991). Plastid chromosomes: structure and evolution. *Molecular Biology of Plastids, 7,* 5–53.
9. Seido, K., Maeda, H., Shiraishi, S. (2000). Determination of the selfing rate in a hinoki (*Chamaecyparis obtusa*) seed orchard by using a chloroplast PCR-SSCP marker. *Silvae Genetica, 49(3),* 165–168.
10. Strauss, S. H., Palmer, J. D., Howe, G. T., Doerksen, A. H. (1988). Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proceedings of the National Academy of Sciences of the United States of America, 85(11),* 3898–3902. DOI 10.1073/pnas.85.11.3898.
11. Wu, C., Wang, Y., Hsu, C., Lin, C., Chaw, S. (2011). Loss of different inverted repeat copies from the chloroplast genomes of Pinaceae and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biology and Evolution, 3(3),* 1284–1295. DOI 10.1093/gbe/evr095.
12. Hao, Z., Cheng, T., Zheng, R., Xu, H., Zhou, Y. et al. (2016). The complete chloroplast genome sequence of a relict conifer *Glyptostrobus pensilis*: Comparative analysis and insights into dynamics of chloroplast genome rearrangement in *Cupressophytes* and Pinaceae. *PLoS One, 11(8),* e161809.
13. Chen, J., Hao, Z., Xu, H., Yang, L., Liu, G. et al. (2015). The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng. *Frontiers in Plant Science, 6,* 447.
14. Chen, C., Zheng, Y., Liu, S., Zhong, Y., Wu, Y. et al. (2017). The complete chloroplast genome of *Cinnamomum camphora* and its comparison with related Lauraceae species. *PeerJ, 5(2),* e3820. DOI 10.7717/peerj.3820.
15. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N. et al. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO Journal, 5(9),* 2043–2049. DOI 10.1002/j.1460-2075.1986.tb04464.x.
16. Lu, L., Li, X., Hao, Z., Yang, L., Zhang, J. et al. (2018). Phylogenetic studies and comparative chloroplast genome analyses elucidate the basal position of halophyte *Nitraria sibirica* (Nitrariaceae) in the Sapindales. *Mitochondrial DNA Part A, 29(5),* 745–755. DOI 10.1080/24701394.2017.1350954.
17. Zheng, W., Chen, J., Hao, Z., Shi, J. (2016). Comparative analysis of the chloroplast genomic information of *Cunninghamia lanceolata* (Lamb.) hook with sibling species from the genera *Cryptomeria* D. Don, Taiwania Hayata, and *Calocedrus* Kurz. *International Journal of Molecular Sciences, 17(7),* 1084. DOI 10.3390/ijms17071084.
18. Chaw, S. M., Chang, C. C., Chen, H. L., Li, W. H. (2004). Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution, 58(4),* 424–441. DOI 10.1007/s00239-003-2564-9.

19. Jakobsson, M., Sall, T., Lind-Hallden, C., Hallden, C. (2010). The evolutionary history of the common chloroplast genome of *Arabidopsis thaliana* and *A. suecica*. *Journal of Evolutionary Biology, 20(1),* 104–121. DOI 10.1111/j.1420-9101.2006.01217.x.
20. Muse, S. V. (2000). Examining rates and patterns of nucleotide substitution in plants. *Plant Molecular Biology, 42 (1),* 25–43. DOI 10.1023/A:1006319803002.
21. Huang, H., Shi, C., Liu, Y., Mao, S. Y., Gao, L. Z. (2014). Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evolutionary Biology, 14(1),* 151. DOI 10.1186/1471-2148-14-151.
22. Shen, Y., Chen, K., Gu, C., Zheng, S., Ma, L. (2018). Comparative and phylogenetic analyses of 26 Magnoliaceae species based on complete chloroplast genome sequences. *Canadian Journal of Forest Research, 48(12),* 1456–1469. DOI 10.1139/cjfr-2018-0296.
23. Wu, C. S., Lin, C. P., Hsu, C. Y., Wang, R. J., Chaw, S. M. (2011). Comparative chloroplast genomes of pinaceae: insights into the mechanism of diversified genomic organizations. *Genome Biology and Evolution, 3,* 309–319. DOI 10.1093/gbe/evr026.
24. Kang, H., Lee, H. O., Lee, I. H., Kim, I. S., Lee, S. et al. (2019). Complete chloroplast genome of *Pinus densiflora* Siebold & Zucc and comparative analysis with five pine trees. *Forests, 10(7),* 600. DOI 10.3390/f10070600.
25. Doyle, J. J., Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin, 19,* 11–15.
26. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M. et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology, 19(5),* 455–477. DOI 10.1089/cmb.2012.0021.
27. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., Pirovano, W. (2010). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics, 27(4),* 578–579. DOI 10.1093/bioinformatics/btq683.
28. Nadalin, F., Vezzi, F., Policriti, A. (2012). GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics, 13(14),* 1–16. DOI 10.1186/1471-2105-13-S14-S8.
29. Liu, C., Shi, L., Zhu, Y., Chen, H., Zhang, J. et al. (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics, 13(1),* 715. DOI 10.1186/1471-2164-13-715.
30. Finn, R. D., Clements, J., Eddy, S. R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research, 39(suppl),* W29–W37. DOI 10.1093/nar/gkr367.
31. Laslett, D., Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research, 32(1),* 11–16. DOI 10.1093/nar/gkh152.
32. Lohse, M., Drechsel, O., Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics, 52(5),* 267–274. DOI 10.1007/s00294-007-0161-y.
33. Li, X., Li, Y., Zang, M., Li, M., Fang, Y. (2018). Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*. *International Journal of Molecular Sciences, 19(8),* 2443. DOI 10.3390/ijms19082443.
34. Mudunuri, S. B., Nagarajaram, H. A. (2007). IMEx: imperfect microsatellite extractor. *Bioinformatics, 23(10),* 1181–1187. DOI 10.1093/bioinformatics/btm097.
35. Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J. et al. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Research, 29(22),* 4633–4642. DOI 10.1093/nar/29.22.4633.
36. Katoh, K., Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution, 30(4),* 772–780. DOI 10.1093/molbev/mst010.
37. Zhang, Z., Li, J., Zhao, X., Wang, J., Wong, G. K. et al. (2006). KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics, Proteomics & Bioinformatics, 4(4),* 259–263. DOI 10.1016/S1672-0229(07)60007-2.

38. Tzeng, Y., Pan, R., Li, W. (2004). Comparison of three methods for estimating rates of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution, 21(12),* 2290–2298. DOI 10.1093/molbev/msh242.
39. Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E. et al. (2011). The variant call format and VCFtools. *Bioinformatics, 27(15),* 2156–2158. DOI 10.1093/bioinformatics/btr330.
40. Darling, A. C., Mau, B., Blattner, F. R., Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research, 14(7),* 1394–1403. DOI 10.1101/gr.2289704.
41. Grant, J. R., Stothard, P. (2008). The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Research, 36(Web Server),* W181–W184. DOI 10.1093/nar/gkn179.
42. Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41(2),* 95–98.
43. Wang, R., Cheng, C., Chang, C., Wu, C., Su, T. et al. (2008). Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology, 8(1),* 36. DOI 10.1186/1471-2148-8-36.
44. Shen, X., Wu, M., Liao, B., Liu, Z., Bai, R. et al. (2017). Complete chloroplast genome sequence and phylogenetic analysis of the medicinal plant *Artemisia annua*. *Molecules, 22(8),* 1330. DOI 10.3390/molecules22081330.
45. Asaf, S., Khan, A. L., Khan, M. A., Shahzad, R., Kang, S. M. et al. (2018). Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS One, 13(3),* e0192966. DOI 10.1371/journal.pone.0192966.
46. Wu, C. S., Chaw, S. M. (2014). Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): evolution towards shorter intergenic spacers. *Plant Biotechnology Journal, 12(3),* 344–353. DOI 10.1111/pbi.12141.
47. Jansen, R. K., Raubeson, L. A., Boore, J. L., Depamphilis, C. W., Chumley, T. W. et al. (2005). Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology, 395,* 348–384.
48. Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J. et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One, 8(2),* e57607. DOI 10.1371/journal.pone.0057607.
49. Kaessmann, H. (2010). Origins, evolution, and phenotypic impact of new genes. *Genome Research, 20(10),* 1313–1326. DOI 10.1101/gr.101386.109.
50. Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M., Rochaix, J. D. (1997). The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO Journal, 16(20),* 6095–6104. DOI 10.1093/emboj/16.20.6095.
51. Li, X., Tan, W., Sun, J., Du, J., Zheng, C. et al. (2019). Comparison of four complete chloroplast genomes of medicinal and ornamental *Meconopsis* species: genome organization and species discrimination. *Scientific Reports, 9(1),* 10567. DOI 10.1038/s41598-019-47008-8.
52. Li, B., Lin, F., Huang, P., Guo, W., Zheng, Y. (2017). Complete chloroplast genome sequence of *Decaisnea* insignis: Genome organization, genomic resources and comparative analysis. *Scientific Reports, 7(1),* 10073. DOI 10.1038/s41598-017-10409-8.
53. Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., Uthaipaisanwong, P., Yoocha, T. et al. (2009). The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Research, 17(1),* 11–22. DOI 10.1093/dnares/dsp025.
54. Qian, J., Song, J., Gao, H., Zhu, Y., Xu, J. et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One, 8(2),* e57607. DOI 10.1371/journal.pone.0057607.
55. Xu, J., Liu, Q., Hu, W., Wang, T., Xue, Q. et al. (2015). Dynamics of chloroplast genomes in green plants. *Genomics, 106(4),* 221–231. DOI 10.1016/j.ygeno.2015.07.004.
56. Morris, L. M., Duvall, M. R. (2010). The chloroplast genome of *Anomochloa marantoidea* (Anomochlooideae; Poaceae) comprises a mixture of grass-like and unique features. *American Journal of Botany, 97(4),* 620–627. DOI 10.3732/ajb.0900226.

57. Hao, D. C., Xiao, P. G., Huang, B. L., Ge, G. B., Yang, L. (2008). Interspecific relationships and origins of Taxaceae and Cephalotaxaceae revealed by partitioned Bayesian analyses of chloroplast and nuclear DNA sequences. *Plant Systematics and Evolution, 276(1),* 89–104. DOI 10.1007/s00606-008-0069-0.
58. Hirao, T., Watanabe, A., Miyamoto, N., Takata, K. (2010). Development and characterization of chloroplast microsatellite markers for *Cryptomeria japonica* D. Don. *Molecular Ecology Resources, 9(1),* 122–124. DOI 10.1111/j.1755-0998.2008.02216.x.
59. Jean-Fran Ois, P., Christian, O., Claude, L., Monique, T. (2005). The chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. *Molecular Biology and Evolution, 22(9),* 1903–1918. DOI 10.1093/molbev/msi182.
60. Pombert, J. F., Lemieux, C., Turmel, M. (2006). The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biology, 4(1),* 3. DOI 10.1186/1741-7007-4-3.
61. Hirao, T., Watanabe, A., Kurita, M., Kondo, T., Takata, K. (2008). Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biology, 8(1),* 70. DOI 10.1186/1471-2229-8-70.
62. Cauz-Santos, L. A., Munhoz, C. F., Rodde, N., Cauet, S., Santos, A. A. et al. (2017). The chloroplast genome of *Passiflora edulis* (Passifloraceae) assembled from long sequence reads: structural organization and phylogenomic studies in Malpighiales. *Frontiers in Plant Science, 8(238),* 334. DOI 10.3389/fpls.2017.00334.
63. Zhou, T., Chen, C., Wei, Y., Chang, Y., Bai, G. et al. (2016). Comparative transcriptome and chloroplast genome analyses of two related *Dipteronia* species. *Frontiers in Plant Science, 7,* 1512.
64. Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L. et al. (2016). Comparative analysis of the complete chloroplast genomes of five *Quercus* species. *Frontiers in Plant Science, 7,* 959.
65. Gao, L., Yi, X., Yang, Y., Su, Y., Wang, T. (2009). Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evolutionary Biology, 9(1),* 130. DOI 10.1186/1471-2148-9-130.
66. Allender, C. J., Allainguillaume, J., Lynn, J., King, G. J. (2007). Simple sequence repeats reveal uneven distribution of genetic diversity in chloroplast genomes of *Brassica oleracea* L. and (n = 9) wild relatives. *Theoretical and Applied Genetics, 114(4),* 609–618. DOI 10.1007/s00122-006-0461-5.
67. Leigh, F. J., Mackay, I., Oliveira, H. R., Gosman, N. E., Horsnell, R. A. et al. (2013). Using diversity of the chloroplast genome to examine evolutionary history of wheat species. *Genetic Resources and Crop Evolution, 60(6),* 1831–1842. DOI 10.1007/s10722-013-9957-4.
68. Provan, J. (2000). Novel chloroplast microsatellites reveal cytoplasmic variation in *Arabidopsis thaliana*. *Molecular Ecology, 9(12),* 2183–2185. DOI 10.1046/j.1365-294X.2000.105316.x.
69. Flannery, M. L., Mitchell, F., Coyne, S., Kavanagh, T. A., Burke, J. I. et al. (2006). Plastid genome characterisation in *Brassica* and Brassicaceae using a new set of nine SSRs. *Theoretical and Applied Genetics, 113(7),* 1221–1231. DOI 10.1007/s00122-006-0377-0.
70. Jones, E., Dupal, M., Dumsday, J., Hughes, L., Forster, J. (2002). An SSR-based genetic linkage map for perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics, 105(4),* 577–584. DOI 10.1007/s00122-002-0907-3.
71. Yi, M., Zhang, L., Lei, L., Cheng, Z., Sun, S. (2020). Analysis of SSR information in transcriptome and development of EST-SSR molecular markers in *Pinus elliottii* Engelm. In *Journal of Nanjing Forestry University (Natural Science Edition),* vol. 44, (in Chinese), 75–83.
72. Rong, H., Huang, B., Zhou, Q., Zhang, W., Xu, L. (2018). The construction of fingerprints and genetic diversity analysis of 61 *Malus crabapple* cultivars based on SSR markers. *Journal of Nanjing Forestry University (Natural Science Edition), 42(3),* 45–50 (in Chinese).
73. Vizoso, P., Undurraga, S. F., Velozo, J. (2019). Chloroplast genome of the soap bark tree *Quillaja saponaria*. *Frontiers in Ecology and Evolution, 7,* 104. DOI 10.3389/fevo.2019.00104.

74. Yang, Y., Dang, Y. Y., Li, Q., Lu, J. J., Li, X. W. et al. (2014). Complete chloroplast genome sequence of poisonous and medicinal plant *Datura stramonium*: organizations and implications for genetic engineering. *PLoS One, 9(11),* e110656. DOI 10.1371/journal.pone.0110656.

75. Asaf, S., Khan, A. L., Khan, M. A., Waqas, M., Kang, S. et al. (2017). Chloroplast genomes of *Arabidopsis halleri* ssp. gemmifera and *Arabidopsis lyrata* ssp. petraea: structures and comparative analysis. *Scientific Reports, 7(1),* 7556. DOI 10.1038/s41598-017-07891-5.

76. Redwan, R. M., Saidin, A., Kumar, S. V. (2015). Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. *BMC Plant Biology, 15(1),* 196. DOI 10.1186/s12870-015-0587-1.

77. Bae, K., Sim, S., Hong, J., Choi, K., Kim, D. et al. (2015). Development of genomic SSR markers and genetic diversity analysis in cultivated radish (*Raphanus sativus* L.). *Horticulture, Environment, and Biotechnology, 56 (2),* 216–224. DOI 10.1007/s13580-015-0089-y.

78. Särkinen, T., George, M. (2013). Predicting plastid marker variation: can complete plastid genomes from closely related species help? *PLoS One, 8(11),* e82266. DOI 10.1371/journal.pone.0082266.

79. Korotkova, N., Nauheimer, L., Ter-Voskanyan, H., Allgaier, M., Borsch, T. (2014). Variability among the most rapidly evolving plastid genomic regions is lineage-specific: implications of pairwise genome comparisons in *Pyrus* (Rosaceae) and other angiosperms for marker choice. *PLoS One, 9(11),* e112998. DOI 10.1371/journal.pone.0112998.

80. Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y. et al. (2015). *Ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports, 5(1),* 8348. DOI 10.1038/srep08348.

81. Parks, M., Liston, A., Cronn, R. (2011). Newly developed primers for complete *ycf1* amplification in *Pinus* (Pinaceae) chloroplasts with possible family-wide utility. *American Journal of Botany, 98(7),* e185–e188. DOI 10.3732/ajb.1100088.

82. Kimura, M. (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Genome, 31 (1),* 24–31. DOI 10.1139/g89-009.

83. Ivanova, Z., Sablok, G., Daskalova, E., Zahmanova, G., Apostolova, E. et al. (2017). Chloroplast genome analysis of resurrection tertiary relict *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Frontiers in Plant Science, 8(402),* 204. DOI 10.3389/fpls.2017.00204.

84. Xu, J., Liu, Q., Hu, W., Wang, T., Xue, Q. et al. (2015). Dynamics of chloroplast genomes in green plants. *Genomics, 106(4),* 221–231. DOI 10.1016/j.ygeno.2015.07.004.

85. Navarro, A., Barton, N. H. (2003). Chromosomal speciation and molecular divergence–accelerated evolution in rearranged chromosomes. *Science, 300(5617),* 321–324. DOI 10.1126/science.1080600.

86. Mentink, R. A., Tsiantis, M. (2015). From limbs to leaves: common themes in evolutionary diversification of organ form. *Frontiers in Genetics, 6,* 284. DOI 10.3389/fgene.2015.00284.

87. Farjon, A. (2005). *Monograph of Cupressaceae and Sciadopitys*. The United Kingdom: Kew Publishing.

88. Farjon, A., Hiep, N. T., Harder, D. K., Loc, P. K., Averyanov, L. (2002). A new genus and species in Cupressaceae (Coniferales) from northern Vietnam, *Xanthocyparis vietnamensis*. *Novon, 12(2),* 179–189. DOI 10.2307/3392953.

89. Quinn, C. J., Price, R. A., Gadek, P. A. (2002). Familial concepts and relationships in the conifer based on *rbcL* and *matK* sequence comparisons. *Kew Bulletin, 57(3),* 513–531. DOI 10.2307/4110984.

90. Stefanović, S., Jager, M., Deutsch, J., Broutin, J., Masselot, M. (1998). Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. *American Journal of Botany, 85(5),* 688–697. DOI 10.2307/2446539.

91. Schulz, C., Stützel, T. (2007). Evolution of taxodiaceous Cupressaceae (Coniferopsida). *Organisms Diversity & Evolution, 7(2),* 124–135. DOI 10.1016/j.ode.2006.03.001.

92. Vuolo, F., Mentink, R. A., Hajheidari, H., Baily, C. D., Filatov, D. A. et al. (2017). Coupled enhancer and coding sequence evolution of a homeobox gene shaped leaf diversity. *Genes & Development, 30(21),* 2370–2375.

93. Zhao, H., Mo, J. X., Hua, H., Guo, Z. H., Xu, J. (2019). Meiosis process and abnormal behavior of pollen mother cells in *Cryptomeria fortunei*. *Journal of Nanjing Forestry University (Natural Science Edition), 43(3),* 45–50 (in Chinese).
94. Quan, W., Zheng, F., Rong, J. (2019). Soil carbon density and C/N distribution of seven forest types in Wuyanling Nature Reserve, Zhejiang Province. *Journal of Nanjing Forestry University (Natural Science Edition), 43(1),* 175–180 (in Chinese).
95. Quan, W., Rong, J., Zheng, F. (2018). Distribution of soil organic carbon among different forest types in Wuyanling Nature Reserve. *Journal of Nanjing Forestry University (Natural Science Edition), 42,* 198–202 (in Chinese).
96. Tsumura, Y., Yoshimura, K., Tomaru, N., Ohba, K. (1995). Molecular phylogeny of conifers using PCR-RFLP analysis of chloroplast genes. *Theoretical and Applied Genetics, 91(8),* 1222–1236. DOI 10.1007/BF00220933.
97. Tsumura, Y., Kimura, M., Nakao, K., Kentaro, U., Tokuko, U. et al. (2020). Effects of the last glacial period on genetic diversity and genetic differentiation in *Cryptomeria japonica* in East Asia. *Tree Genetics & Genomes, 16(1),* 19. DOI 10.1007/s11295-019-1411-0.

Appendix

Appendix A: Lengths of exons and introns with introns in the *C. fortunei* chloroplast genome

Gene	Exon I (bp)	Intron I (bp)	Exon II (bp)	Intron II (bp)	Exon III (bp)
atpF	151 $^{+}$	698 $^{+}$	410 $^{+}$		
ndhA	559 $^{-}$	758 $^{-}$	548 $^{-}$		
ndhB	723 $^{+}$	690 $^{+}$	756 $^{+}$		
rpl2	403 $^{-}$	677 $^{-}$	431 $^{-}$		
rpoC1	441 $^{+}$	709 $^{+}$	1656 $^{+}$		
rps12	331^{+}	585 $^{+}$	62 $^{+}$		
rps16	31 $^{-}$	599 $^{-}$	113 $^{-}$		
ycf2	1,920 $^{+}$	33 $^{+}$	4965 $^{+}$		
ycf3	124 $^{+}$	699 $^{+}$	230 $^{+}$	697 $^{+}$	156 $^{+}$
trnI-AAU	30 $^{+}$	509 $^{+}$	61 $^{+}$		
trnS-CGA	32 $^{-}$	766 $^{-}$	61 $^{-}$		
trnL-UAA	35 $^{+}$	473 $^{+}$	50 $^{+}$		
trnA-UGC	39 $^{-}$	770 $^{-}$	40 $^{-}$		
trnS-UUA	38 $^{-}$	2401 $^{-}$	55 $^{-}$		
trnG-UUC	34 $^{-}$	856 $^{-}$	45 $^{-}$		

Note: $^{+}$, genes transcribed in the clockwise direction; $^{-}$, genes transcribed in the counterclockwise direction.

Appendix B: Codon-anticodon recognition patterns and codon usage for the *C. fortunei* chloroplast genome

Amino acid	Codon	No.	RSCU	Amino acid	Codon	No.	RSCU	Amino acid	Codon	No.	RSCU
Ala	GCA	427	1.17	Glu	GAA	1123	1.57	Ser	AGC	108	0.38
	GCC	175	0.48		GAG	311	0.43		AGU	358	1.28
	GCG	139	0.38	His	CAC	114	0.44		UCA	333	1.19
	GCU	723	1.98		CAU	404	1.56		UCC	235	0.84
Asn	AAC	251	0.44	Ile	AUA	702	1.00		UCG	137	0.49
	AAU	901	1.56		AUC	359	0.51		UCU	514	1.83
Asp	GAC	206	0.38		AUU	1035	1.48	Ter	UAA	42	1.54
	GAU	868	1.62	Leu	CUA	393	0.87		UAG	17	0.62
Arg	AGA	450	1.91		CUC	162	0.36		UGA	23	0.84
	AGG	134	0.57		CUG	154	0.34	Thr	ACA	362	1.20
	CGA	294	1.25		CUU	541	1.20		ACC	202	0.67
	CGC	117	0.50		UUA	894	1.99		ACG	136	0.45
	CGG	82	0.35		UUG	552	1.23		ACU	508	1.68
	CGU	337	1.43	Lys	AAA	1200	1.53	Trp	UGG	433	1
Cys	UGC	71	0.52		AAG	367	0.47	Tyr	UAC	195	0.42
	UGU	201	1.48	Met	AUG	581	1		UAU	734	1.58
Gly	GGA	639	1.64	Phe	UUC	427	0.60	Val	GUA	502	1.51
	GGC	158	0.40		UUU	999	1.40		GUC	163	0.49
	GGG	208	0.53	Pro	CCA	294	1.12		GUG	198	0.60
	GGU	558	1.43		CCC	190	0.73		GUU	465	1.40
Gln	CAA	708	1.58		CCG	120	0.46				
	CAG	190	0.42		CCU	442	1.69				

Appendix C: Long repeat structures in the *C. fortunei* chloroplast genome

ID	Repeat Start 1	Type	Size (bp)	Repeat Start 2	Mismatch (bp)	E-Value	Gene
1	19,692	D	30	19,716	0	4.22×10^{-9}	*rps18*
2	23,999	D	31	24,172	0	1.06×10^{-9}	
3	86,691	D	31	86,693	0	1.06×10^{-9}	
4	107,755	P	31	121,523	0	1.06×10^{-9}	
5	5836	P	32	5836	0	2.64×10^{-10}	
6	23,649	P	32	23,685	0	2.64×10^{-10}	
7	86,691	P	32	86,691	0	2.64×10^{-10}	
8	86,692	P	32	86,692	0	2.64×10^{-10}	
9	23,930	D	33	24,170	0	6.60×10^{-11}	

(Continued)

Appendix C (continued).

ID	Repeat Start 1	Type	Size (bp)	Repeat Start 2	Mismatch (bp)	E-Value	Gene
10	47,143	P	34	47,143	0	1.65×10^{-11}	*chlB*
11	23,894	D	35	24,099	0	4.12×10^{-12}	
12	23,892	D	37	23,963	0	2.58×10^{-13}	
13	23,892	D	37	24,030	0	2.58×10^{-14}	
14	63,091	P	38	63,091	0	6.44×10^{-14}	
15	24,316	D	40	130,998	0	4.03×10^{-15}	
16	10,952	P	42	10,952	0	2.52×10^{-16}	
17	131,160	P	44	131,160	0	1.57×10^{-17}	
18	121,741	D	47	121,855	0	2.46×10^{-19}	
19	23,965	D	50	24,099	0	3.84×10^{-21}	
20	121,637	D	51	121,875	0	9.60×10^{-22}	
21	107,682	P	57	121,569	0	2.34×10^{-25}	
22	24,032	D	60	24,099	0	3.66×10^{-27}	
23	40,566	P	71	113,308	0	8.73×10^{-34}	*trnA-CAU*
24	121,679	D	71	121,855	0	8.73×10^{-34}	
25	23,932	D	83	23,999	0	5.21×10^{-41}	
26	121,637	D	89	121,699	0	1.27×10^{-44}	
27	128,649	D	102	128,715	0	1.89×10^{-52}	*ycf1*
28	76,434	D	113	76,455	0	4.52×10^{-59}	
29	115,880	D	195	115,913	0	1.93×10^{-108}	*ycf2*
30	47,937	P	275	85,655	0	1.32×10^{-156}	

Note: D, forward repeats; P, palindromic repeats (including inverted and complementary sequences)

Appendix D: Quantitative analysis of long repeat structures (LRSs) in chloroplast genomes in *C. fortunei* and related species, sorted by LRS length

Lengh of LRS (bp)	*C. fortunei*	*C. japonica*	*C. japonica cv.*	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. flousiana*	*T. distichum*
30	1	3	1	2	0	0	2	1	1	2
31	3	2	1	0	0	0	0	8	7	0
32	4	2	2	3	1	2	3	2	2	1
33	1	0	0	2	4	0	0	3	3	0
34	1	3	1	1	11	1	4	3	3	1
35	1	0	0	1	1	1	4	1	1	1
36	0	0	0	1	0	0	2	2	2	1
37	2	2	2	0	0	0	3	0	0	0
38	1	1	1	1	0	1	2	0	0	1
39	0	0	0	1	1	0	0	1	0	0

Appendix D (continued).

Lengh of LRS (bp)	*C. fortunei*	*C. japonica*	*C. japonica cv.*	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. flousiana*	*T. distichum*
40	1	1	1	2	0	1	0	1	1	0
41	0	0	0	2	0	0	4	0	0	1
42	1	2	1	0	0	1	1	1	2	1
43	0	0	1	2	1	0	0	0	0	0
44	1	1	2	1	0	1	1	0	0	1
45	0	1	0	3	0	0	0	0	0	0
46	0	2	0	0	2	0	1	0	2	0
47	1	0	3	0	0	0	3	1	1	1
48	0	0	0	0	1	0	0	0	0	0
49	0	0	0	0	0	0	0	2	1	1
50	1	0	0	0	0	0	0	0	0	0
51	1	1	1	0	0	0	1	1	1	0
52	0	1	0	0	1	0	0	0	0	0
53	0	0	0	0	2	1	0	0	0	0
54	0	0	0	0	1	0	0	0	0	2
55	0	0	0	0	1	0	0	0	0	0
57	1	1	1	0	0	0	0	0	0	0
58	0	0	0	0	0	0	1	1	1	0
60	1	0	0	0	2	0	1	0	0	0
61	0	0	0	1	0	0	3	0	0	1
63	0	0	0	0	2	0	0	0	0	0
65	0	0	0	1	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0	0	1
69	0	0	0	0	0	0	0	1	0	0
70	0	0	0	0	0	0	1	0	0	0
71	2	2	2	1	0	0	0	1	1	0
72	0	0	0	0	0	0	0	1	2	1
73	0	0	0	1	0	0	0	0	0	0
74	0	0	0	1	0	0	0	0	0	0
76	0	0	1	0	0	0	0	0	0	2
79	0	0	0	0	0	0	0	1	1	0
81	0	0	0	0	0	0	1	0	0	0
83	1	0	0	1	0	0	0	0	0	1
86	0	0	0	0	0	1	0	0	0	0
88	0	0	0	0	0	0	1	0	0	0
89	1	0	1	1	0	0	0	0	0	0
95	0	0	0	1	0	1	0	1	0	0
98	0	0	0	1	0	0	0	0	0	0
99	0	0	0	0	1	0	0	0	0	0
102	1	0	0	0	0	0	0	0	0	0

(Continued)

Appendix D (continued).

Lengh of LRS (bp)	*C. fortunei*	*C. japonica*	*C. japonica cv.*	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. flousiana*	*T. distichum*
105	0	0	1	0	0	0	0	0	0	0
109	0	0	1	0	0	0	0	0	0	0
111	0	0	0	1	0	0	0	0	0	0
113	1	1	1	0	0	0	0	0	0	0
114	0	0	0	1	0	0	0	0	0	0
116	0	0	0	0	1	0	0	0	0	0
117	0	0	0	0	0	0	0	0	1	0
119	0	0	0	0	0	0	1	0	0	0
121	0	0	1	1	0	0	0	0	0	0
122	0	0	0	1	0	0	0	0	0	0
124	0	0	0	0	0	0	0	1	1	0
129	0	1	0	0	0	0	0	0	0	0
134	0	0	0	0	0	0	1	0	0	0
136	0	0	0	1	0	0	0	0	0	0
137	0	0	0	0	0	0	0	1	1	0
144	0	0	0	0	0	0	1	0	0	0
146	0	0	0	0	1	0	0	0	0	0
159	0	0	0	0	0	0	0	1	1	0
162	0	0	1	0	0	0	0	0	0	0
170	0	0	0	0	1	0	0	0	0	0
173	0	0	0	0	0	0	1	0	0	0
195	1	0	0	0	0	0	0	0	0	0
202	0	0	0	0	0	0	0	1	1	0
234	0	0	1	0	0	0	0	0	0	0
271	0	0	0	0	0	0	0	1	1	0
273	0	0	0	0	0	0	1	0	0	0
275	1	1	0	0	0	0	0	0	0	0

Appendix E: Simple sequence repeats in the *C. fortunei* chloroplast genome

ID	Repeat Motif	Length (bp)	Start	End	Gene	ID	Repeat Motif	Length (bp)	Start	End	Gene
1	(T)8	8	1690	1,697	*cemA*	108	(T)8	8	64,831	64,838	
2	(A)10	10	2949	2,958	*ycf4*	109	(A)11	11	65,052	65,062	
3	(A)13	13	3134	3,146		110	(T)9	9	65,086	65,094	
4	(AT)10	20	3150	3,169		111	(T)10	10	65,325	65,334	
5	(T)8	8	3188	3,195		112	(T)8	8	65,346	65,353	
6	(T)9	9	3673	3,681		113	(TA)7	14	65,552	65,565	

Appendix E (continued).

ID	Repeat Motif	Length (bp)	Start	End	Gene	ID	Repeat Motif	Length (bp)	Start	End	Gene
7	(T)10	10	5865	5,874		114	(A)11	11	66,882	66,892	*rpoB*
8	(TAT)3	9	7889	7,897		115	(A)11	11	67,284	67,294	
9	(TGC)3	9	10,634	10,642	*rps11*	116	(AGA)3	9	69,015	69,023	
10	(A)10	10	10,939	10,948		117	(G)8	8	71,145	71,152	
11	(AT)8	16	10,966	10,981		118	(GAA)3	9	71,507	71515	*rpoC1*
12	(TCA)4	12	11,260	11,271	*infA*	119	(AAT)3	9	71,701	71,709	
13	(A)9	9	11,621	11,629		120	(ATG)3	9	72,512	72,520	
14	(T)8	8	11,680	11,687		121	(CTT)3	9	74,448	74,456	*rpoC2*
15	(T)9	9	11,918	11,926	*rps8*	122	(A)8	8	74,537	74,544	
16	(T)11	11	12,218	12,228		123	(A)9	9	74,842	74,850	
17	(T)8	8	12,336	12,343		124	(A)8	8	75,794	75,801	
18	(T)8	8	14,156	14,163		125	(T)10	10	76,388	76,397	
19	(T)9	9	14,628	14,636	*rps3*	126	(A)8	8	76,723	76,730	*rps2*
20	(AT)5	10	14,683	14,692		127	(A)11	11	77,083	77,093	
21	(AGT)4	12	14,943	14,954		128	(ATG)3	9	77,336	77,344	
22	(T)15	15	15,469	15,483	*rps19*	129	(GAA)3	9	77,923	77,931	*atpI*
23	(TCT)3	9	15,673	15,681		130	(TA)5	10	78,707	78,716	
24	(A)9	9	15,789	15,797		131	(AT)6	12	78,720	78,731	
25	(AT)9	18	16,346	16,363		132	(A)9	9	78,793	78,801	
26	(AT)5	10	16,365	16,374		133	(G)9	9	79,301	79,309	
27	(CCT)3	9	17,153	17,161	*rpl2*	134	(TAGA)3	12	79,377	79,388	
28	(TTA)3	9	18,506	18,514		135	(T)9	9	79,391	79,403	
29	(TA)6	12	18,677	18,688		136	(A)9	9	80,277	80,285	
30	(T)9	9	18,689	18,697		137	(AAT)3	9	82,832	82,840	
31	(AAT)3	9	19,202	19,210	*rpl20*	138	(A)12	12	82,855	82,866	
32	(GTT)3	9	19,496	19,504	*rps18*	139	(T)10	10	83,425	83,434	
33	(T)10	10	19,815	19,824		140	(T)8	8	83,753	83,760	
34	(T)9	9	20,328	20,336		141	(A)8	8	84,250	84,257	
35	(T)9	9	20,428	20,436		142	(A)8	8	85,250	85,257	*psbK*
36	(T)20	20	20,696	20,715		143	(A)8	8	85,817	85,824	
37	(T)9	9	20,847	20,855		144	(CAA)3	9	86,516	86,524	
38	(T)9	9	20,980	20,988		145	(T)8	8	86,607	86,614	
39	(A)8	8	21,001	21,008		146	(TA)16	32	86,692	86,723	
40	(A)11	11	21,105	21,115		147	(ATA)4	12	86,987	86,998	

(Continued)

Appendix E (continued).

ID	Repeat Motif	Length (bp)	Start	End	Gene	ID	Repeat Motif	Length (bp)	Start	End	Gene
41	(T)9	9	21,407	21,415		148	(TAT)3	9	87,823	87,831	*ndhD*
42	(A)11	11	22,544	22,554		149	(C)8	8	88,201	88,208	
43	(T)9	9	23,525	23,533	*psbJ*	150	(TCC)3	9	88,472	88,480	
44	(T)10	10	23,758	23,767		151	(CAG)3	9	89,518	89,526	*ndhE*
45	(T)11	11	23,931	23,941		152	(A)9	9	90,629	90,637	
46	(T)9	9	24,000	24,008		153	(GAA)3	9	91,324	91,332	
47	(T)9	9	24,067	24,075		154	(TTC)3	9	91,728	91,736	*ndhA*
48	(T)9	9	24,134	24,142		155	(T)8	8	92,146	92,153	
49	(T)14	14	24,168	24,181		156	(T)13	13	92,279	92,291	
50	(A)12	12	24,260	24,271		157	(AGC)3	9	92,705	92,713	*ndhA*
51	(T)8	8	25,470	25,477		158	(A)9	9	92,803	92,811	
52	(A)8	8	25,543	25,550		159	(A)8	8	93,369	93,376	*ndhH*
53	(TTC)3	9	26,673	26,681	*accD*	160	(A)13	13	94,760	94,772	
54	(TTC)3	9	27,033	27,041		161	(T)12	12	95,625	95,636	
55	(T)8	8	27,185	27,192		162	(TTTC)3	12	96,400	96,411	*ndhF*
56	(T)10	10	28,153	28,162		163	(AT)5	10	96,682	96,691	
57	(TA)8	16	28,274	28,289		164	(A)11	11	98,658	98,668	
58	(A)12	12	28,340	28,351		165	(A)10	10	99,097	99,106	
59	(T)17	17	28,479	28,495		166	(A)10	10	99,242	99,251	
60	(TATG)3	12	30,451	30,462		167	(CAC)3	9	99,538	99,546	
61	(A)10	10	30,526	30,535		168	(TA)5	10	99,904	99,913	
62	(T)9	9	33,781	33,789		169	(CTAC)3	12	101,953	101,964	*rrn23*
63	(T)16	16	34,058	34,073		170	(CTT)3	9	104,861	104,869	
64	(A)9	9	34,521	34,529		171	(TCT)3	9	105,097	105,105	
65	(T)8	8	34,908	34,915		172	(TG)5	10	105,706	105,715	
66	(GTT)3	9	35,562	35,570	*ndhK*	173	(T)9	19	105,715	105,724	
67	(A)12	12	36,744	36,755		174	(A)8	8	105,791	105,798	
68	(TCTA)3	12	36,821	36,832		175	(AT)8	16	107,976	107,991	
69	(T)11	11	39,656	39,666		176	(TAT)3	9	108,354	108,362	
70	(TTA)3	9	40,045	40,053		177	(A)13	13	108,560	108,572	
71	(ACAT)3	12	40,077	40,088		178	(A)15	15	108,821	108,835	
72	(T)8	8	40,192	40,199		179	(A)8	8	108,888	108,895	
73	(TTA)5	15	40,548	40,562		180	(AT)8	16	108,972	108,987	
74	(TTC)3	9	41,126	41,134	*psbA*	181	(ATG)3	9	109,441	109,449	

Appendix E (continued).

ID	Repeat Motif	Length (bp)	Start	End	Gene	ID	Repeat Motif	Length (bp)	Start	End	Gene
	(GCT)3	9	41,397	41,405		182	(T)13	13	109,694	109,706	
76	(ATA)3	9	42,127	42,135		183	(A)15	15	111,960	111,974	
77	(TTGA)3	12	42,810	42,821	*matK*	184	(TAT)3	9	112,755	112,763	*ndhB*
78	(ATC)3	9	42,876	42,884		185	(TAT)3	9	113,710	113,718	
79	(A)8	8	43,998	44,005		186	(CT)5	10	113,753	113,762	
80	(T)9	9	44,543	44,551		187	(A)8	8	114,707	114,714	*ycf2*
81	(AAAG)3	12	44,626	44,637		188	(A)8	8	116,540	116,547	
82	(C)10	10	45,723	45,732		189	(A)8	8	119,097	119,104	
83	(T)14	14	45,979	45,992		190	(CTC)3	9	119,123	119,131	
84	(A)10	10	46,069	46,078		191	(AGA)3	9	119,901	119,909	
85	(AT)6	12	46,322	46,333		192	(ATA)3	9	121,667	121,675	
86	(T)8	8	48,044	48,051		193	(A)9	9	121,693	121,701	
87	(A)9	9	48,774	48,782	*rps4*	194	(ATA)3	9	121,729	121,737	
88	(A)8	8	49,188	49,195		195	(A)9	9	121,755	121,763	
89	(A)10	10	50,275	50,284		196	(ATA)3	9	121,791	121,799	
90	(A)10	10	51,408	51,417		197	(A)9	9	121,869	121,877	
91	(A)8	8	51,626	51,633		198	(ATA)3	9	121,905	121,913	
92	(T)8	8	52,729	52,736		199	(A)11	11	121,928	121,938	
93	(CAT)3	9	55,884	55,892	*psaB*	200	(TTTA)3	12	122,273	122,284	
94	(TCA)3	9	56,180	56,188		201	(A)10	10	123,209	123,218	*ycf1*
95	(TTA)3	9	57,722	57,730		202	(A)9	9	123,678	123,686	
96	(A)8	8	58,211	58,218		203	(T)8	8	124,244	124,251	
97	(AT)10	20	58,708	58,727		204	(GAA)3	9	124,438	124,446	
98	(TAC)3	9	61,905	61,913		205	(AAG)3	9	126,701	126,709	
99	(ATT)3	9	62,514	62,522		206	(A)8	8	127,248	127,255	
100	(A)10	10	62,621	62,630		207	(AAG)3	9	128,018	128,026	
101	(T)9	9	62,703	62,711		208	(AT)5	10	129,724	129,733	
102	(AT)5	10	62,857	62,866		209	(AT)5	10	129,736	129,745	
103	(T)8	8	63,607	63,614		210	(A)9	9	130,854	130,862	
104	(TAAA)3	12	63,806	63,817		211	(GAA)3	9	130,990	130,998	
105	(T)10	10	64,085	64,094		212	(AT)9	18	131,174	131,191	
106	(T)8	8	64,363	64,370		213	(CAAT)3	12	131,449	131,460	
107	(TAT)3	9	64,380	64,388							

Appendix F: Comparison of simple sequence repeat (SSR) distributions in chloroplast (cp) genomes between *C. fortunei* and related species

	C. fortunei	*C. japonica*	*C. japonica cv.*	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. flousiana*	*T. distichum*
A	60	62	63	72	53	58	43	62	61	69
AAAG	1	1	1	0	0	0	0	1	1	0
AAAGA	0	0	0	0	0	0	0	1	1	0
AAAT	0	0	0	0	0	0	0	0	0	1
AAC	0	0	0	3	3	2	1	2	2	1
AAG	2	2	2	1	1	0	3	2	2	0
AAT	3	3	3	2	4	3	4	4	4	2
ACAT	1	1	1	0	0	0	0	0	0	0
ACG	0	0	0	5	0	0	0	0	0	0
AG	0	0	0	1	1	0	0	1	1	0
AGA	2	2	2	2	1	2	3	3	3	3
AGC	1	1	1	1	1	1	1	1	1	1
AGGT	0	0	0	0	1	1	1	0	0	0
AGT	1	1	1	1	1	1	0	0	0	0
AT	15	15	15	14	14	7	11	7	7	11
ATA	6	7	8	3	4	3	1	2	2	3
ATAA	0	0	0	1	0	1	0	1	1	0
ATATA	0	0	0	1	0	0	0	0	0	0
ATC	1	1	1	2	0	1	1	1	1	1
ATCA	0	0	0	0	1	0	0	0	0	0
ATCT	0	0	0	1	0	0	0	0	0	0
ATG	3	3	3	3	2	2	1	2	2	4
ATGG	0	0	0	0	0	1	1	0	0	0
ATGT	0	0	0	0	0	0	1	0	0	0
ATT	1	1	1	4	2	6	3	5	5	3
ATTA	0	0	0	0	0	0	0	1	1	0
ATTG	0	0	0	0	0	0	0	0	0	1
ATTT	0	0	0	0	1	0	1	0	0	0
C	2	2	2	2	1	2	4	0	0	1
CA	0	0	0	1	0	0	0	0	0	0
CAA	1	1	1	2	0	0	2	1	1	1
CAAT	1	1	1	0	0	0	0	0	0	0
CAC	1	1	1	0	0	0	0	0	0	1
CAG	1	1	1	1	0	0	1	1	1	1
CAT	1	1	1	0	2	1	3	1	1	1
CATA	0	0	0	0	1	1	1	0	0	0
CCT	1	1	1	0	1	1	1	1	1	0
CT	1	1	1	0	0	0	0	0	0	0
CTAA	0	0	0	0	0	0	0	0	0	1
CTAC	1	1	1	1	0	0	0	1	1	1
CTC	1	1	1	1	0	0	0	1	1	1

Appendix F (continued).

	C. fortunei	*C. japonica*	*C. japonica cv.*	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. flousiana*	*T. distichum*
CTT	2	2	2	2	1	2	2	3	3	2
CTTT	0	0	0	0	1	0	0	0	0	0
G	2	2	2	2	1	4	3	2	2	0
GA	0	0	0	0	0	0	0	0	0	1
GAA	5	5	5	5	7	5	3	6	6	4
GAAA	0	0	0	0	1	1	1	0	0	0
GAG	0	0	0	1	1	2	2	1	1	1
GAT	0	0	0	0	2	1	1	0	0	0
GCA	0	0	0	1	0	0	0	0	0	1
GCT	1	1	1	1	1	1	1	1	1	1
GGC	0	0	0	1	0	0	0	0	0	0
GTA	0	0	0	0	0	0	1	0	0	0
GTG	0	0	0	1	1	1	0	1	1	0
GTT	2	2	2	1	1	1	1	1	1	2
T	58	57	58	51	56	57	62	42	42	47
TA	6	6	6	7	4	9	5	12	12	10
TAA	0	0	0	1	2	0	0	5	5	2
TAAA	1	1	1	0	1	0	0	0	0	0
TAAG	0	0	0	0	0	1	1	0	0	0
TAAT	0	0	0	0	0	0	0	1	1	0
TAC	1	1	1	0	1	1	0	0	0	2
TACTA	0	0	0	0	0	0	0	1	1	0
TAG	0	0	0	1	0	3	3	0	0	0
TAGA	1	1	1	0	0	0	0	0	0	0
TAGTA	0	0	0	0	0	0	0	1	1	0
TAT	6	6	6	5	5	1	2	7	7	6
TATG	1	1	1	0	0	0	0	1	1	1
TATT	0	0	0	0	0	0	1	0	0	1
TC	0	0	0	0	1	0	0	0	0	0
TCA	2	2	2	2	2	2	0	2	2	2
TCAA	0	0	0	0	0	1	0	0	0	0
TCC	1	1	1	1	0	0	1	0	0	0
TCCA	0	0	0	1	0	0	0	1	1	0
TCT	2	2	2	0	1	1	2	0	0	0
TCTA	1	1	1	0	0	0	0	0	0	0
TG	1	1	1	0	0	0	0	1	1	1
TGA	0	0	0	0	0	2	1	0	0	0
TGC	1	1	1	0	2	2	1	0	0	0
TTA	4	4	4	3	0	2	1	3	3	2
TTAT	0	0	0	1	0	0	0	1	1	0

(Continued)

Appendix F (continued).

	C. fortunei	*C. japonica*	*C. japonica cv.*	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. flousiana*	*T. distichum*
TTC	4	4	4	8	4	5	6	4	4	5
TTG	0	0	0	1	1	2	1	1	1	0
TTGA	1	1	1	1	0	0	0	1	1	1
TTTA	1	1	1	0	0	0	0	0	0	0
TTTAT	0	0	0	0	0	1	0	0	0	0
TTTC	1	1	1	1	0	0	0	1	1	1
TTTTTC	0	0	0	0	0	0	0	1	1	0

Appendix G: Total length and AT content of simple sequence repeats in Cupressaceae chloroplast genomes

Species	Total length (bp)	AT content (bp)	AT content (%)
C. fortunei	383	322	84.07
C. japonica	387	326	84.24
C. japonica cv. Wogon Hort	392	332	84.69
C. lanceolata	404	328	81.19
G. pensilis	341	287	84.16
M. glyptostroboides	357	293	82.07
S. sempervirens	338	272	80.47
T. cryptomerioides	395	336	85.06
T. flousiana	394	334	84.77
T. distichum	354	298	84.18
Total	3745	3128	83.52

Appendix H: Ka/Ks ratios of nine chloroplast genomes compared with the *C. fortunei* chloroplast genome

ID	Gene	*C. japonica*	*C. japonica cv.* Wogon Hort	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. distichum*	*T. flousiana*
1	*accD*	0	0	0.903	1.080	0.611		0.906	1.134	0.906
2	*atpA*	0	0	0.228	0.252	0.171	0.169	0.192	0.110	0.192
3	*atpB*	0	0	0.101	0.069	0.094	0.125	0.277	0.191	0.277
4	*atpE*	0	0	0.311	0.428	0.315	0.303	0.434	0.413	0.434
5	*atpF*	0	0	0.787	0.979	0.932	0.596	0.826	1.275	0.826
6	*atpH*	0	0	0	0	0	0	0	0	0
7	*atpI*	0	0	0.321	0.728	0.464	0.581	0.374	0.835	0.374
8	*ccsA*	0	0	0.316	0.151	0.355	0.353	0.374	0.092	0.374
9	*cemA*	0	0	0.393	0.569	0.611	0.709	0.409	0.472	0.409
10	*chlB*	0	0	0.127	0.505	0.177	0.151	0.139	0.153	0.139
11	*chlL*	0	0	0.046	0.126	0.110	0.051	0.015	0.066	0.015

Appendix H (continued).

ID	Gene	*C. japonica*	*C. japonica cv.* Wogon Hort	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. distichum*	*T. flousiana*
12	*chlN*	0	0	0.220	0.373	0.196	0.277	0.281	0.584	0.281
13	*clpP*				0.431	0.364	0.331	0.644	0.374	0.644
14	*infA*	0	0		0.234	0.419		1.097	0	1.097
15	*matK*	0		0.422	0.542	0.486	0.476	0.363	0.352	0.363
16	*ndhA*	0.230	0.230	0.165	0.173	0.151		0.255	0.229	0.255
17	*ndhB*	0	0	0.236	0.056	0.184	0.164	0.220	0.079	0.220
18	*ndhC*	0	0	0.268	0.201	0.094	0.094	0.177	0.201	0.177
19	*ndhD*	0	0	0.183	0.100	0.159	0.210	0.231	0.171	0.231
20	*ndhE*	0	0	0.094	0.309	0.157	0.136	0.121	0.231	0.121
21	*ndhF*	0	0	0.344	0.431	0.258	0.232	0.251	0.262	0.251
22	*ndhG*	0	0	0.272	0.439	0.368	0.460	0.458	0.770	0.458
23	*ndhH*	0	0	0.168	0.107	0.051	0.085	0.146	0.180	0.146
24	*ndhI*	0	0	0	0.170	0.021	0	0.029	0	0.029
25	*ndhJ*	0	0	0.148	0.279	0.271	0.365	0.179	0.287	0.179
26	*ndhK*	0	0	0.248	0.660	0.340	0.397	0.397	0.441	0.397
27	*petA*	0	0	0.124	0.177	0.108	0.129	0.109	0	0.109
28	*petB*	2.230	2.230	0.116	0.273	0.130	0.036	0.080	0.179	0.080
29	*petD*	0	0	0.173	1.297	0.364	0.193	0.197	0.668	0.197
30	*petG*	0	0	0	0	0	0	0	0	0
31	*petL*	0	0	0.450	0	0	0	1.308	0	1.308
32	*petN*	0	0	0	0	0	0	0	0	0
33	*psaA*	0	0	0.179	0.106	0.140	0.162	0.160	0.091	0.160
34	*psaB*	0	0	0.194	0	0.167	0.144	0.146	0.024	0.146
35	*psaC*	0	0	0	0	0	0	0	0	0
36	*psaI*	0	0	0.494	0.797	1.214	1.214	1.001	0.348	1.001
37	*psaJ*	0	0	0.736	0	0.349	0.495	0.359	0	0.359
38	*psaM*	0	0		0	2.442	2.442		0	
39	*psbA*	0	0	0	0	0	0	0.078	0	0.078
40	*psbB*	0	0	0.045	0	0.039	0.054	0.033	0	0.033
41	*psbC*	0	0	0.039	0	0.017	0	0	0	0
42	*psbD*	0	0	0.027	0.050	0.035	0.016	0.019	0.041	0.019
43	*psbE*	0	0	0.114	0.114	0.114	0.080	0.062	0	0.062
44	*psbH*	0	0	0.375	0.250	0.209	0.258	0.259	0	0.259
45	*psbI*	0	0	0	0	0.179	0	0	0	0
46	*psbJ*	0	0	0.294	0.271	0.132	0.231	0.294	0.271	0.294
47	*psbK*	0	0	0.283	0	0.472	0.177	0.153	0	0.153
48	*psbL*	0	0	0	0	0	0	0	0	0
49	*psbM*	0	0	1.211	0	1.052	1.052	1.092	0	1.092
50	*psbN*	0	0	0	0	0	0	0	0	0
51	*psbT*	0	0	0	0	0.094	0.113	0	0	0
52	*rbcL*	0	0	0.108	0.170	0.058	0.066	0.068	0.241	0.068

(Continued)

Appendix H (continued).

ID	Gene	*C. japonica*	*C. japonica cv.* Wogon Hort	*C. lanceolata*	*G. pensilis*	*M. glyptostroboides*	*S. sempervirens*	*T. cryptomerioides*	*T. distichum*	*T. flousiana*
53	*rpl14*	0	0	0.067	0	0.142	0.310	0.216	0	0.216
54	*rpl2*	0	0.866	0.166	0.275	0.210	0.196	0.196	0.253	0.196
55	*rpl20*	0	0	0.299	0.067	0.322	0.392	0.573	0.064	0.573
56	*rpl22*	0	0	1.163	0	1.019		1.423	0	1.423
57	*rpl23*	0	0	0.214	1.149	0.019	0	0.666	0	0.666
58	*rpl32*	0	0	0.512	0.322	0.221	0.307		0.114	
59	*rpl33*	0	0	0.156	0	0.153	0.107	0.182	0	0.182
60	*rpl36*	0	0	0.097	0	0	0	0.203	0	0.203
61	*rpoA*	0	0	0.325	0.125	0.374	0.325	0.531	0.309	0.529
62	*rpoB*	0	0	0.604	0.231	0.455	0.468	0.459	0.173	0.459
63	*rpoC1*	0		0.622	0.717	0.594	0.739	0.693	0.651	0.693
64	*rpoC2*	0	0	0.531	0.756	0.484	0.565	0.563	0.706	0.580
65	*rps11*	0	0	0.563	0.179	0.376	0.419	0.498	0.209	0.498
66	*rps12*	0.866	0	0.645	0.806	0.746	0.811	0.546	0.716	0.546
67	*rps14*	0	0	0.187	0.298	0.198	0.204	0.727	0.298	0.727
68	*rps15*	0	0	0.212	0.175	0.360	0.281	0.307	0.106	0.307
69	*rps16*	0	0	1.076	0.546			0.811	0.615	0.811
70	*rps18*	0	0	0.850	0.500	0.369		0.987	0.732	0.987
71	*rps19*	0	0	0.831	0.382	0.524	0.410	0.260	0.382	0.260
72	*rps2*	0	0	0.225	0.229	0.073	0.144	0.517	0.286	0.517
73	*rps3*	0	0	0.375	0.126	0.287		0.678	0.068	0.678
74	*rps4*	0	0	0.287	0	0.216	0.332	0.338	0	0.338
75	*rps7*	0	0	0.350	0	0.272	0.323	0.386	0	0.386
76	*rps8*	0	0	0.905	0.193	0.344	0.271	0.584	0.311	0.584
77	*ycf1*	0	0	0.324	0.339	0.290	0.318	0.377	0.325	0.377
78	*ycf2*	0	0	0.657	0.645	0.546	0.660	0.749		0.745
79	*ycf3*	0	0	0.044	0	0.075	0	0.072	0.174	0.072
80	*ycf4*	0	0	0.089	0.290	0.352	0.218	0.133	0.236	0.133