




## ORIGINAL RESEARCH

# Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*

Jingyao Ping<sup>1</sup>  | Peipei Feng<sup>1</sup> | Jinye Li<sup>1</sup> | Rongjing Zhang<sup>1</sup> | Yingjuan Su<sup>2,3</sup>  | Ting Wang<sup>1</sup> <sup>1</sup>College of Life Sciences, South China Agricultural University, Guangzhou, China<sup>2</sup>School of Life Sciences, Sun Yat-sen University, Guangzhou, China<sup>3</sup>Research Institute of Sun Yat-sen University in Shenzhen, Shenzhen, China**Correspondence**Yingjuan Su, Xingang West Road, Guangzhou, Guangdong, IA 510275, China.  
Email: suyj@mail.sysu.edu.cnTing Wang, Wushan Road, Guangzhou, Guangdong, IA 510640, China.  
Email: tingwang@scau.edu.cn**Funding information**

This work was funded by the National Natural Science Foundation of China (31670200, 31770587, 31872670, and 32071781).

**Abstract**

Chloroplast genome sequences have been used to understand evolutionary events and to infer efficiently phylogenetic relationships. *Callitropsis funebris* (Cupressaceae) is an endemic species in China. Its phylogenetic position is controversial due to morphological characters similar to those of *Cupressus*, *Callitropsis*, and *Chamaecyparis*. This study used next-generation sequencing technology to sequence the complete chloroplast genome of *Ca. funebris* and then constructed the phylogenetic relationship between *Ca. funebris* and its related species based on a variety of data sets and methods. Simple sequence repeats (SSRs) and adaptive evolution analysis were also conducted. Our results showed that the monophyletic branch consisting of *Ca. funebris* and *Cupressus tonkinensis* is a sister to *Cupressus*, while *Callitropsis* is not monophyletic; *Ca. nootkatensis* and *Ca. vietnamensis* are nested in turn at the base of the monophyletic group *Hesperocyparis*. The statistical results of SSRs supported the closest relationship between *Ca. funebris* and *Cupressus*. By performing adaptive evolution analysis under the phylogenetic background of Cupressales, the Branch model detected three genes and the Site model detected 10 genes under positive selection; and the Branch-Site model uncovered that *rpoA* has experienced positive selection in the *Ca. funebris* branch. Molecular analysis from the chloroplast genome highly supported that *Ca. funebris* is at the base of *Cupressus*. Of note, SSR features were found to be able to shed some light on phylogenetic relationships. In short, this chloroplast genomic study has provided new insights into the phylogeny of *Ca. funebris* and revealed multiple chloroplast genes possibly undergoing adaptive evolution.

**KEYWORDS**adaptive evolution, *Callitropsis funebris*, chloroplast genome, phylogeny, SSRs

## 1 | INTRODUCTION

Chloroplast is a unique organelle of green plants and has an independent genetic system. With the development of sequencing

technology, chloroplast genome data can be obtained relatively easily, and the massive emergence of chloroplast genome data provides sufficient resources for the study of chloroplast genomics (Tonti-Filippini et al., 2017). The size of a typical terrestrial plant chloroplast genome

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

is generally 120–160 kb, with a quadripartite structure involving two inverted repeat (IR) sequences that separate the rest of the genome into large and small single-copy regions (Kwon et al., 2020; Wicke et al., 2011). In particular, conifers are one of the few groups that have lost the canonical IR region (Guo et al., 2014; Kwon et al., 2020; Li et al., 2016; Ping et al., 2020). Compared with nuclear genome and mitochondrion genome, chloroplast genome is characterized by small genome size, high copy number, conservative structure, and moderate nucleotide evolution rate, which has been widely used to understand evolutionary events and to infer efficiently phylogenetic relationships (Barrett et al., 2016; Pacheco et al., 2019; Saina et al., 2018). Similarly, the conservation of chloroplast genes and chloroplast genome structure, gene loss and pseudonymization, gene transfer events, plastid rearrangement, and RNA editing sites make it possible to characterize specific lineages (Bock, 2017; De Santana Lopes et al., 2019; Martin et al., 2014; Vieira et al., 2016).

The repeat sequence, gene sequence, gene content, and GC content of the chloroplast genome can provide more information for phylogeny. Simple sequence repeats (SSRs), also known as microsatellite, are a piece of DNA composed of basic units of 1 to 6 nucleotides that are repeated many times, and it has polymorphism rate at the species level (Cavalier-Smith, 2002). They are widely distributed in different locations in the genome and are generally <200 bp in length. SSR molecular markers are widely used because of their advantages of codominance, high information content, wide coverage, and easy operation. The SSRs of chloroplast genomes of plants contain rich information on genetic variation and have been widely used in crop species identification, genetic diversity, and relationship studies (Chmielewski et al., 2015) and have a broad prospect of application in plant genetics and taxonomy (Jiménez, 2010).

Chloroplasts, as organelles for plant photosynthesis, are closely related to the adaptation of plants to various environments. Some genes that play a key role in photosynthesis may undergo adaptive evolution along with species radiation (Hasegawa et al., 2009; Zhang et al., 2018). Studies have reported that *clpP* (Erixon & Oxelman, 2008), *rbL* (Kapralov & Filatov, 2007), *matK* (Hao et al., 2010), and other chloroplast genes have undergone positive selection. Analysis of adaptive evolution at the genetic level and identification of the key functional sites and amino acid structure can provide evidence for exploring the mysteries of plant evolution, as well as a deeper understanding of genetic structure and function variations (Nei & Kumar, 2000).

Gymnosperms originated more than 300 million years ago and have about 860 species. It includes four of the five main lineages of seed plants: cycads, ginkgos, gnetophytes, and conifers. There are 227 gymnosperm species in China, covering 8 families and 37 genera (Wang et al., 2015). Due to the large genome size, high heterozygosity, and long generation time of populations, understanding the evolution of gymnosperms still faces great difficulties in the era of genomics (Wang & Ran, 2014). At the beginning of the 21st century, *Cupressus* was a hot topic in the conifer systematics, and various suggestions have been proposed to subdivide it into multiple genera (Adams et al., 2009; de Laubenfels, 2009; Farjon et al., 2002; Little, 2006; Mao et al., 2010; Terry et al., 2012). Now cypresses are divided into

two groups: Old World cypresses (only *Cupressus*) and New World cypresses (*Cupressus*, *Callitropsis*, and *Hesperocyparis*). There are about 25 species of Old World cypresses, among which *Callitropsis funebris* is endemic to China. It is widely distributed, with rapid growth, strong adaptability, and other characteristics. As a precious wood species, it also has medicinal value and is of very important economic values.

The scaly branchlets of *Ca. funebris* are flat, drooping, with smaller cones and a fewer number of seeds per scale (only 5–6). These characteristics are similar to those of *Chamaecyparis* and *Callitropsis*; but the branchlets of *Ca. funebris* are drooping and homomorphic, unlike those of *Chamaecyparis*, which are flattened and heteromorphic. The cones of *Ca. funebris* mature in the second year and have 3–4 pairs of seed scales with 4 pollen sacs in each scale that are similar to those of *Cupressus*. Zheng and Fu (1978) believed that the main morphological traits of *Ca. funebris* were consistent with those of *Cupressus*, and it was more natural to put it into *Cupressus*. Similarly, Rushforth et al. (2003) divided *Ca. funebris* into Old World cypresses based on molecular trait analysis. The relationship between *Ca. funebris* and *Chamaecyparis* has also been controversial (Farjon, 2005; Gadek & Quinn, 1987; Jagel & Stutzel, 2001). More recently, De Laubenfels et al. (2012) further tried to solve the problems existing in New World cypresses and Old World cypresses according to morphological characteristics, believing that *Callitropsis nootkatensis*, *Xanthocyparis vietnamensis*, and *Cupressus funebris* belong to *Callitropsis*. However, morphological analysis is often limited in the construction of phylogenetic relationships, and the application of molecular data can provide more information for phylogenetic development.

Due to the unique characteristics of *Ca. funebris* and its controversial phylogenetic location, we start from the chloroplast genome to explore its phylogenetic location and further understand related evolutionary events. So, in this study, the complete chloroplast genome of *Ca. funebris* was sequenced using next-generation sequencing technology. Phylogenetic relationship of Cupressales was constructed based on a variety of data sets and methods. We also analyzed and compared the chloroplast genome characteristics of cypresses (focus on SSRs) and carried out adaptive evolution analysis under the phylogenetic background of Cupressales.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling

Fresh leaves from *Ca. funebris* were sampled from the campus of South China Agricultural University (E113°35', N23°15'). The leaves were wrapped in tinfoil and stored in the liquid nitrogen at –80°C for later use.

### 2.2 | DNA extraction and sequencing assembly

The chloroplast genomic DNA was extracted using a new DNasecure plant genomic DNA extraction kit (TIANGEN), and the DNA library

(300 bp) was constructed after fragmentation. Illumina HiSeq2500 platform was used to carry out two-end sequencing with 150 bp reading length. In order to ensure the quality, the original data must be filtered to remove sequences with joints and low quality. Trimmomatic v0.32 (Bolger et al., 2014) was used to filter data to obtain clean reads. Clean data were spliced and assembled using velvet v1.2.03 (Zerbino & Birney, 2008). DOGMA software (Wyman et al., 2004) was used to predict gene and gene function annotation, and rpsblast v2.2.30+ (Altschul et al., 1997) was used to predict gene COG functional information. Chloroplast genome mapping was performed by OGDRAW v1.3 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) (Greiner et al., 2019). Sequence was submitted through the BankIt platform (<https://www.ncbi.nlm.nih.gov/WebSub/>).

### 2.3 | Construction of phylogenetic tree

The chloroplast genome sequences of 31 species of Cupressales were downloaded from the NCBI database (Table 1), and three sequence datasets were constructed, respectively. Datasets were as follows: (a) 32 chloroplast genome sequences, (b) 32 sequences of *rbcl* and *matK* (*rbcl* + *matK*), and (c) chloroplast genome complete sequences of 14 cypresses and *Juniperus monosperma*. The Clustal W module in MEGA X (Kumar et al., 2018) was used for sequence alignment and correction, and Jmodeltest 2.1.7 was used to predict the

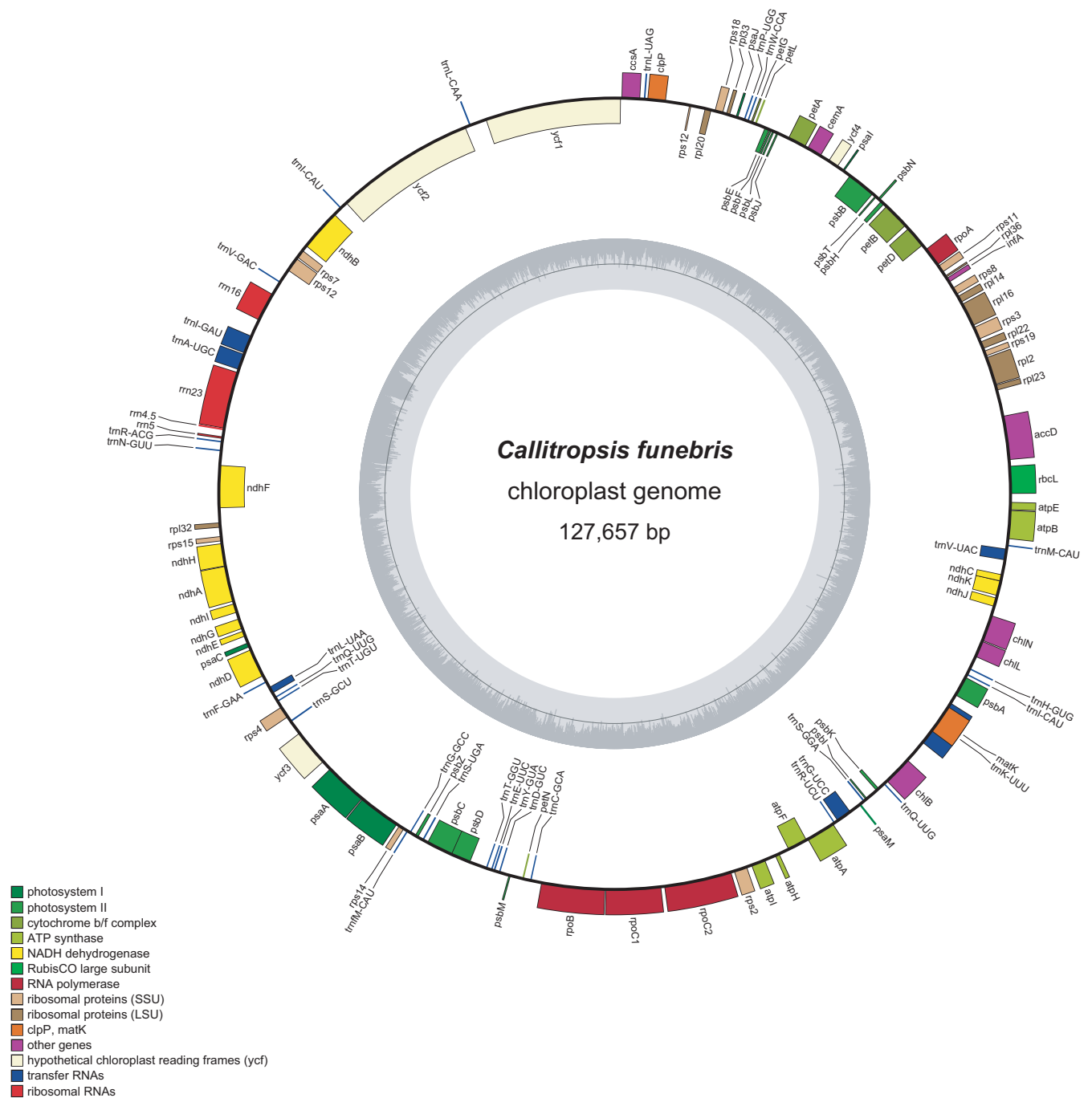
optimal nucleotide substitution model. MEGA X was used to construct a neighbor-joining (NJ) tree. Maximum parsimony (MP) tree was constructed by using PAUP4.0 (Swofford, 2002). RaxmlGUI2 (Stamatakis, 2014) was applied to construct maximum likelihood (ML) tree, and MrBayes3.2.6 (Huelsenbeck & Ronquist, 2001) was used to conduct bayesian inference (BI) tree. Finally, FigTree v1.4.4 was used to draw the phylogenetic tree. The phylogenetic tree for evolutionary analysis was obtained after comparison and correction based on the comprehensive analysis of each tree file and the results of previous studies at the genus level (Gadek et al., 2000; Hao et al., 2016; Lu et al., 2014; Qu, Jin et al., 2017; Qu, Wu et al., 2017).

### 2.4 | Analysis of chloroplast genome structure characteristics

The files containing the complete sequence of the chloroplast genome were imported into OGDRAW and Geneious Prime 2020.1.2 (Kearse et al., 2012), and the genomic structure information was generated. MISA (Microsatellite Identification Tool) online website (<https://webblast.ipk-gatersleben.de/misa/index.php?action=1>) was employed to predict SSRs. The parameters were set to default settings; SSR motif length corresponding to the minimum number of repetitions is 1–10, 2–6, 3–5, 4–5, 5–5, 6–5. When the distance between two microsatellites is <100 bp, 2 microsatellites were considered to form a composite microsatellite (Beier et al., 2017).

**TABLE 1** The information of sample species

Species name	Genbank accession number	Species name	Genbank accession number
Cupressaceae			
<i>Cupressus chengiana</i>	NC_034788	<i>Juniperus monosperma</i>	NC_024022
<i>Cu. gigantea</i>	NC_028155	<i>J. squamata</i>	NC_044076
<i>Cu. jiangeensis</i>	NC_036939	<i>Chamaecyparis formosensis</i>	NC_034943
<i>Cu. sempervirens</i>	NC_026296	<i>Ch. hodginsii</i>	NC_036996
<i>Cu. tonkinensis</i>	NC_039562	<i>Thuja sutchuenensis</i>	NC_042176
<i>Cu. torulosa</i>	NC_039563	<i>Thujopsis dolabrata</i>	KX832628
<i>Callitropsis funebris</i>	MT227813	<i>Callitris rhomboidea</i>	NC_034940
<i>Ca. nootkatensis</i>	KP099642	<i>Taxodium distichum</i>	NC_034941
<i>Ca. vietnamensis</i>	KX832629	<i>Glyptostrobos pensilis</i>	NC_031354
<i>Hesperocyparis glabra</i>	KX832624	<i>Cryptomeria japonica</i>	NC_010548
<i>Hesperocyparis lindleyi</i>	NC_039566	<i>Sequoia sempervirens</i>	NC_030372
<i>H. lusitanica</i>	MH121051	<i>Metasequoia glyptostroboides</i>	NC_027423
<i>H. benthamii</i>	NC_039565	<i>Taiwania flousiana</i>	NC_021441
<i>H. arizonica</i>	NC_039564	<i>Cunninghamia lanceolata</i>	NC_021437
<i>Calocedrus formosana</i>	NC_023121	<i>Platycladus orientalis</i>	KX832626
Taxaceae			
<i>Amentotaxus argotaenia</i>	NC_027581		
Sciadopityaceae			
<i>Sciadopitys verticillata</i>	NC_029734		



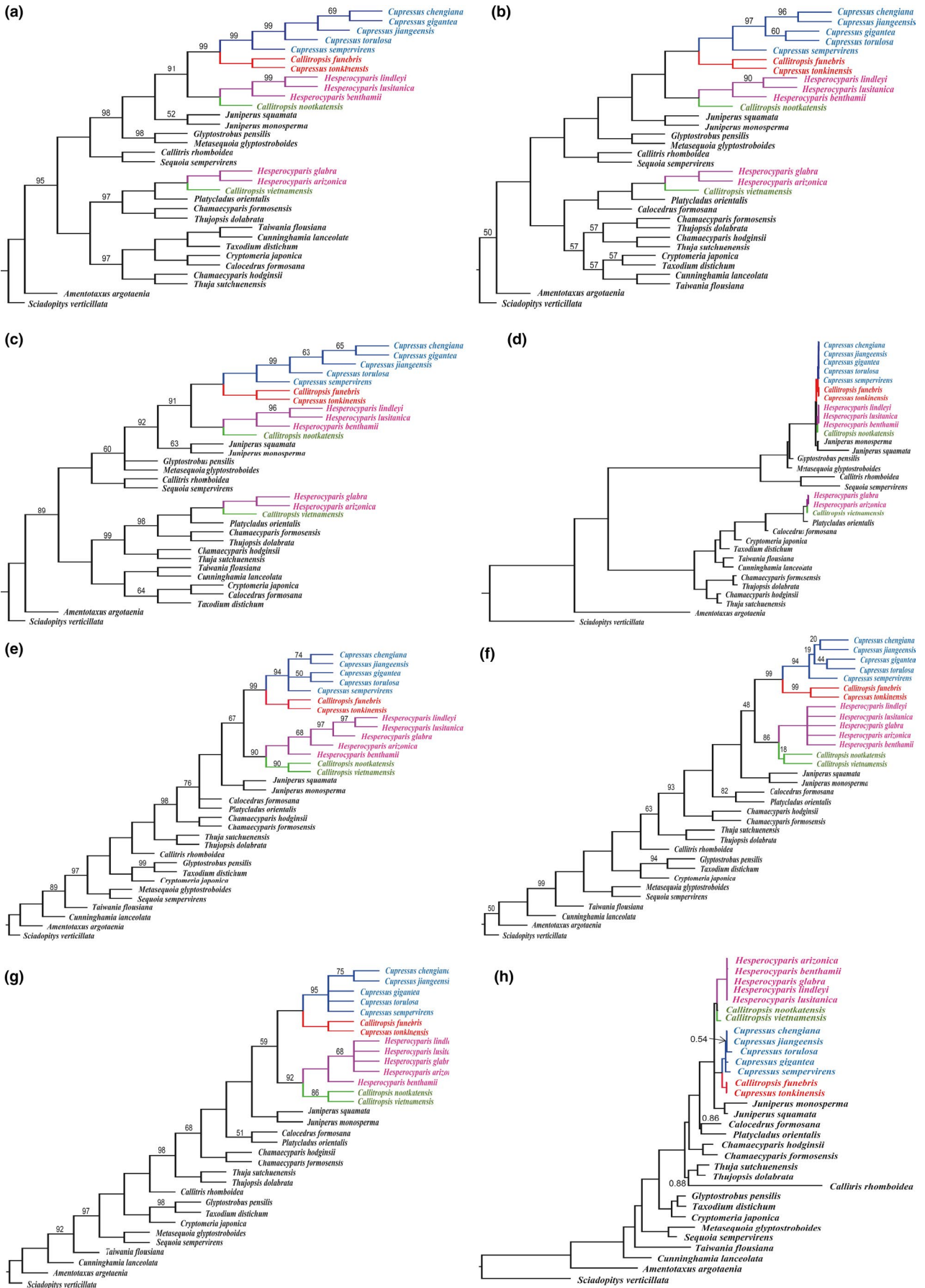
**FIGURE 1** Gene map of the *Callitropsis funebris* chloroplast genome

### 2.5 | Adaptive evolutionary analysis

Geneious Prime 2020.1.2 was used to screen the shared genes of the chloroplast genomes of 32 species, and to extract the coding sequences of the genes. The sequences were aligned and corrected using the

ClustalW (Codons) module in MEGA X. Sequence files in.fasta/.fas format were converted to .PML format through DAMBE 7.2.1 (Xia, 2017). Using the codeml program of software PAML 4.9 (Yang, 2007), adaptive evolutionary analysis of common protein-coding genes was performed based on the ML method in the phylogenetic background of Cupressales.

**FIGURE 2** Phylogenetic relationship of Cupressales was constructed by four methods and two data sets. (a–d) based on the complete chloroplast genome. (e–h) based on the series data of *rbcl* and *matK* (*rbcl* + *matK*). (a–d) and (e, f) correspond to NJ tree, MP tree, ML tree, and Bayes tree, respectively. The number on the branch is the bootstrap support value. Blue branch: *Cupressus* (except *Cu. tonkinensis*). Red branch: *Ca. funebris* and *Cu. tonkinensis*. Purple branches: *Hesperocyparis*. Green branch: *Ca. nootkatensis* and *Ca. vietnamensis*



Three models were used: (a) Branch Model, the variable  $\omega$  model between branches (Yang, 1998) allows the  $\omega$  value to change in different branches. M0 (One-ratio) assumes that all evolutionary branches have the same  $\omega$  value. F (Free-ratio) assumes that each branch has different  $\omega$  values. The M0 and F likelihood ratio test can further determine the positive selection in the branch. Model2 (two-ratio) assumes that the values of foreground branch and background branch are different. Here, because the cypresses is the research object, the cypresses branch and the ancestor branch of *Cupressus tonkinensis* and *Ca. funebris* were set as the foreground branch, respectively, to examine whether there is any difference in the values of foreground branch and background branch. (b) Site Model, variable  $\omega$  model between sites (Yang et al., 2000) assumes that different sites have different values of  $\omega$  and there is no difference in different branches of the phylogenetic tree. It mainly includes four pairs of nested models, among them M1a (near neutral) and M2a (positive selection) were for detecting positive selection sites. Site models can be used by BEB (Bayes Empirical Bayes) to identify positively selected sites (Yang et al., 2005). (c) Branch-Site Model (Yang & dos Reis, 2010). It analyses genes that accept the F model in the Branch model. The alternative hypothesis with cypresses or the ancestor branch of *Cu. tonkinensis* and *Ca. funebris* as the foreground branch and the corresponding null hypothesis with  $\omega = 1$  were, respectively, tested for likelihood ratio, and the positive selection sites were screened by BEB.

### 3 | RESULTS

#### 3.1 | Chloroplast genome structure of *Ca. funebris*

Raw data of 3.66G were generated, including 12,206,029 raw reads. Standard reads were 150 bp, and 3.37G clean data were obtained after trimming. The complete genome sequence of chloroplast was obtained by assembly, with the total length of 127,657 bp and GC content of 34.7%; and there is no IR sequence in this genome (Figure 1). A total of 115 genes were encoded, including 82 protein-coding genes, 31 tRNA genes (*trnI*-CAU and *trnQ*-UUG are double-copy genes), and 4 rRNA genes. Twelve genes contained one intron, and two genes (*rps12*, *ycf3*) contained two introns. The Genbank accession number is MT227813.

BLAST search in NCBI showed that *Ca. funebris* had the highest homology with *Cu. tonkinensis* (Query Cover: 100%, Per.ident: 99.87%). The sequence similarity between *Cu. tonkinensis* and *Ca. funebris* is extremely high, and they are sister groups in the phylogenetic tree. Meanwhile, in NCBI *Cu. tonkinensis* Silba is treated as homotypic synonym as *Ca. funebris* (*Ca. funebris*) subsp. *tonkinensis*. Therefore, in the subsequent analysis and discussion, *Cu. tonkinensis* is not regarded as *Cupressus* species, but in the same category with *Ca. funebris*.

#### 3.2 | Phylogenetic relationship of Cupressaceae

Taking *Sciadopitys verticillata* as the outgroup, phylogenetic analysis based on the complete sequence of the chloroplast genome showed that the relationship among Cupressaceae was chaotic (Figure 2a–d). Like the *Chamaecyparis*, the species branches of *Hesperocyparis* were not clustered together. And the support among the branches was low. Phylogenetic analysis based on *rbcl* + *matK* showed clearly the relationship of Cupressaceae (Figure 2e–h). Among them, cypresses form a monophyletic branch and the monophyletic branch of *Ca. nootkatensis* and *Ca. vietnamensis* is sister group to *Hesperocyparis*. Also, branch relationships among other genera have higher support. It should be noted that there is a paraphyletic relationship between *Cupressus* and *Hesperocyparis*; the relationship between the related species is difficult to determine. All phylogenetic trees consistently showed that *Cu. tonkinensis* and *Ca. funebris* are sister branches to each other, and the monophyletic branch formed by the two are sister group to the *Cupressus*.

#### 3.3 | Chloroplast genome structure of cypresses

Chloroplast genome size of cypresses is 127,005 bp–129,150 bp, GC content is 34.6%–34.7%, and the number of genes, protein-coding genes, and tRNA coding genes is highly consistent, 119, 82, and 33, respectively. The difference is that the genome size of *Cupressus* is between 128,151 and 129,150 bp, and that of *Hesperocyparis* is between 127,005 and 127,158 bp. Numerically, *Ca. funebris* is closer to *Cupressus*. *Ca. nootkatensis* and *Ca. vietnamensis* are closer to *Hesperocyparis* (Table 2).

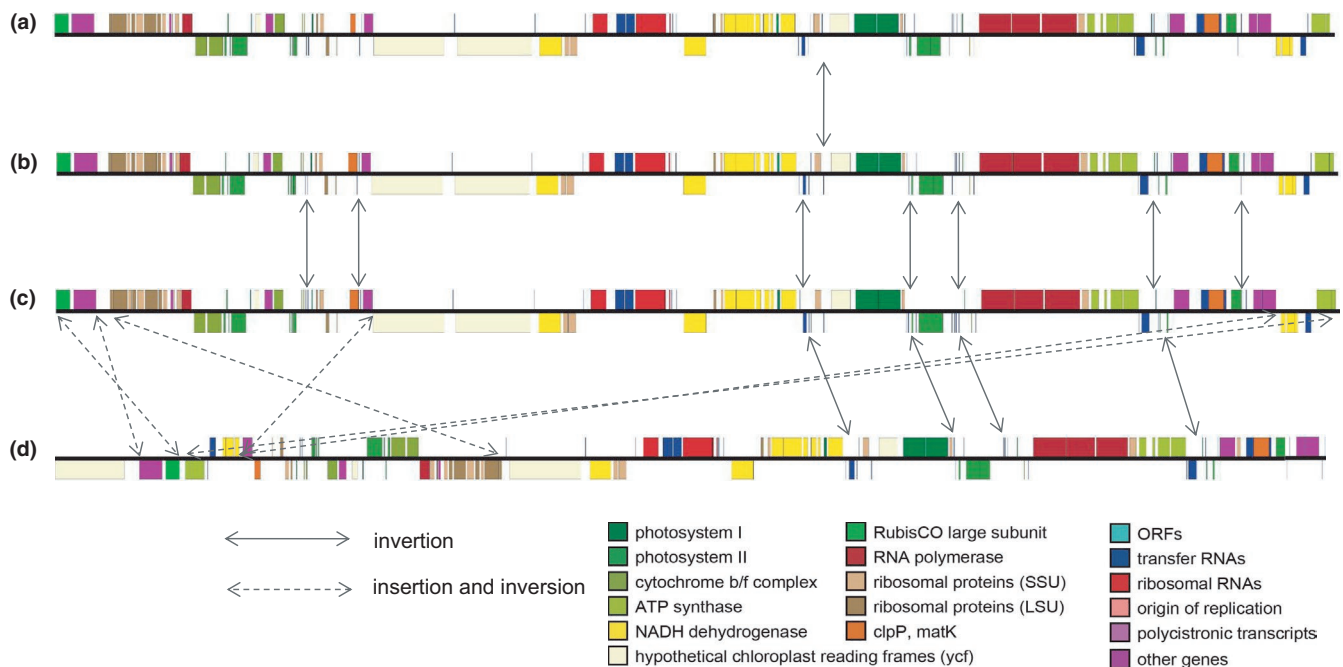
Genes of the cypress species are in the same order and are divided into three types according to the location of the genes (Figure 3). Type A includes *Cupressus* and *Ca. nootkatensis*, type B includes *Hesperocyparis* and *Ca. vietnamensis*, and type C includes *Ca. funebris* and *Cu. tonkinensis*. Only one gene is inverted between type A and type B, and there are 11 genes that are inverted between type C and type A, as well as between type C and type B. Of note, type D contains *Chamaecyparis*. At least 3 insertion-inversion events and inversion of multiple genes occur between type D and type C.

#### 3.4 | Phylogenetic relationship of cypresses

In order to further determine the phylogenetic relationship of cypresses, the complete chloroplast genome sequences of 14 cypresses and *J. monosperma* were analyzed. The genes of all 15 species are in the same order on the genome, with *J. monosperma* as the outgroup. The results showed that the phylogenetic relationship obtained by the four methods was consistent. The monophyletic branch formed by *Ca. funebris* and *Cu. tonkinensis* is the sister group of the remaining *Cupressus*. *Ca. nootkatensis* and *Ca. vietnamensis* were in turn nested at the base of the monophyletic group *Hesperocyparis*. These branch relationships are strongly supported (Figure 4a).

Species name	Genome size/bp	GC content	Genes	Protein-coding genes	tRNA
<i>Cu. chengiana</i>	128,151	34.70%	119	82	33
<i>Cu. gigantea</i>	128,244	34.70%	119	82	33
<i>Cu. jiangeensis</i>	128,286	34.70%	119	82	33
<i>Cu. sempervirens</i>	129,150	34.60%	119	82	33
<i>Cu. torulosa</i>	128,322	34.60%	119	82	33
<i>Cu. tonkinensis</i>	127,835	34.70%	119	82	33
<i>Ca. funebris</i>	127,657	34.70%	119	82	33
<i>Ca. nootkatensis</i>	127,150	34.70%	119	82	33
<i>Ca. vietnamensis</i>	127,479	34.70%	119	82	33
<i>H. glabra</i>	127,064	34.70%	119	82	33
<i>H. lindleyi</i>	127,005	34.70%	119	82	33
<i>H. lusitanica</i>	127,113	34.70%	119	82	33
<i>H. benthamii</i>	127,007	34.70%	119	82	33
<i>H. arizonica</i>	127,158	34.70%	119	82	33
<i>Ch. formosensis</i>	127,211	35.00%	121	85	32
<i>Ch. hodginsii</i>	127,777	35.00%	120	83	33

**TABLE 2** Genome structure characteristics of cypresses and *Chamaecyparis*



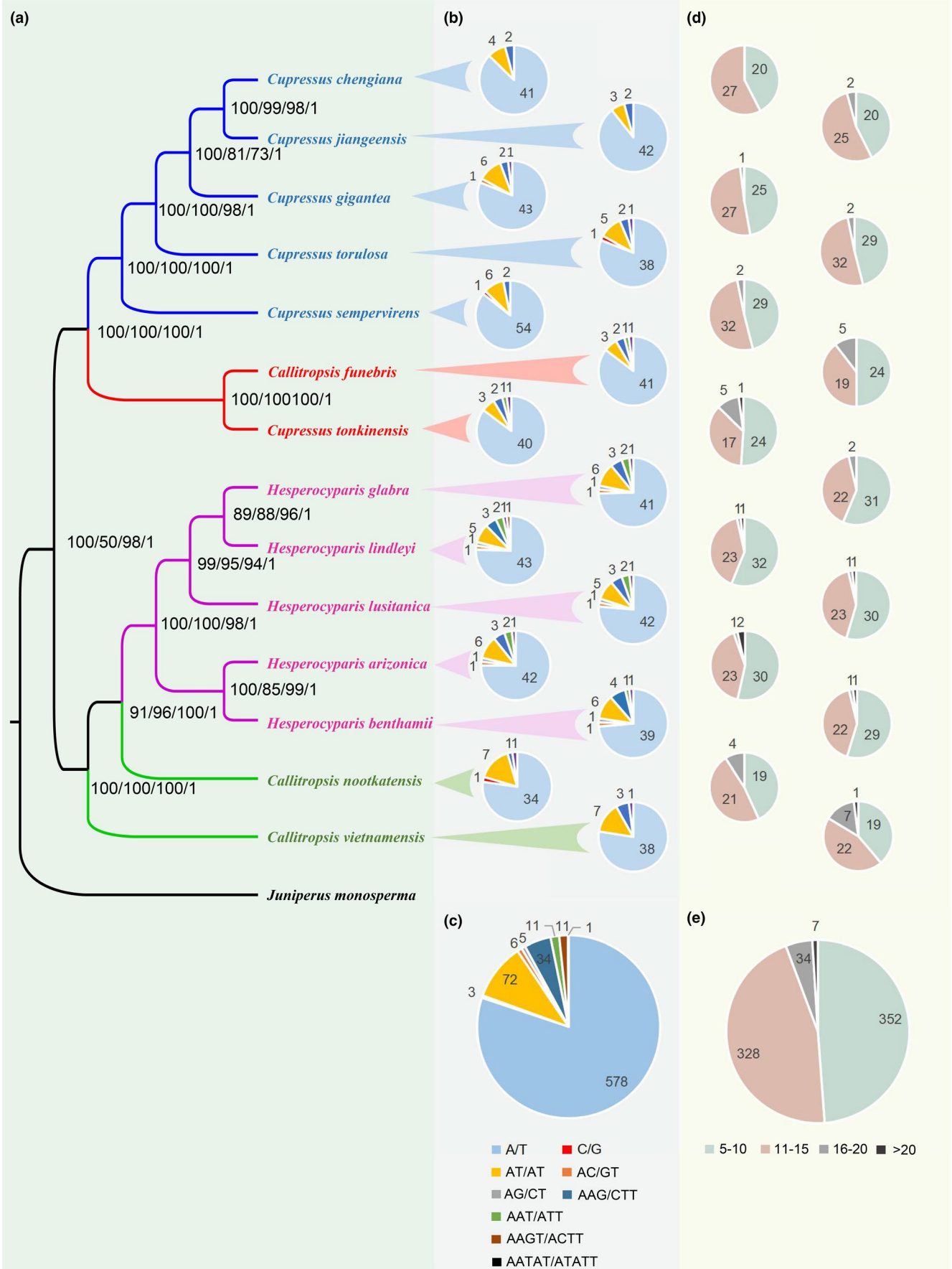
**FIGURE 3** Gene arrangement of chloroplast genomes of cypresses and *Chamaecyparis*. (a) *Cu. chengiana*, *Cu. gigantea*, *Cu. jiangeensis*, *Cu. sempervirens*, and *Ca. nootkatensis*. (b) *Cu. torulosa*, *H. lusitanica*, *H. glabra*, *H. lindleyi*, *H. benthamii*, *H. arizonica*, *Ca. vietnamensis*. (c) *Ca. funebris*, *Cu. tonkinensis*. (d) *Ch. formosensis*, *Ch. hodginsii*

### 3.5 | | SSR analysis of cypresses

All cypresses detected a total of 721 SSRs, including a total of 5 repeat types and 8 types of repeat motifs (Figure 4c and Appendix S1). The number of mononucleotide repeats is the largest, totaling 581 (80.6%), of which A/T repeat motif accounts for 99.4%. Dinucleotide repeats contain 3 types of repeat motifs (AT/AT, AC/GT, AG/CT), in total 83, of which AT/AT accounts for 86.75%. Trinucleotide repeats

have two kinds of repeat motifs, a total of 45, mainly AAG/CTT (75.56%). Both tetranucleotide repeats and pentanucleotide repeats contain only one kind of repeat motif. Pentanucleotide repeats were only found in *H. lindleyi*. Overall, SSRs are dominated by AT base (A/T, AT/AT, AAT/ATT, AATAT/ATATT), accounting for 91.82% of the total.

Different species have different types and contents of motifs. Of more obvious features, AG/CT exists only in the *Hesperocyparis*.





**FIGURE 4** Phylogenetic relationship of cypresses and SSRs analysis. (a) With the complete sequence of chloroplast genome, the phylogenetic relationship of cypresses was constructed based on NJ, MP, ML, and BAYES, and the number of node locations, respectively, represents the branch support corresponding to the four methods. (b) the number and type of the motifs of each species. (c) the number and type of motifs of all species. (d) distribution of the number of repetitions in each species. (e) distribution of the number of repetitions for all species

AAT/ATT is found in *Ca. funebriis*, *Cu. tonkinensis* and *Hesperocyparis*, not in *Callitropsis* and *Cupressus*. *Hesperocyparis* has the most abundant types of motifs, except for *H. lindleyi*, which contains eight types of repeat motifs, and the remaining four species contain seven types, respectively (Figure 4b and Appendix S1). In terms of content, due to the scarcity of G/C motif, the proportion of A/T motif is close to that of mononucleotide, and the proportions of mononucleotide repeats were counted in each species (Figure 5): *Cupressus* (81%–89%), *Hesperocyparis* (74%–76%), *Ca. funebriis*, *Ca. nootkatensis* and *Ca. vietnamensis*: 85%, 80%, 78%. The proportion of trinucleotide repeats in various species is: *Cupressus* (3.2%–4.3%), *Ca. funebriis* and *Cu. tonkinensis* (6.3%, 6.4%), *Hesperocyparis* (8.9%–9.4%), *Ca. nootkatensis* and *Ca. vietnamensis* (2.3%, 6.1%).

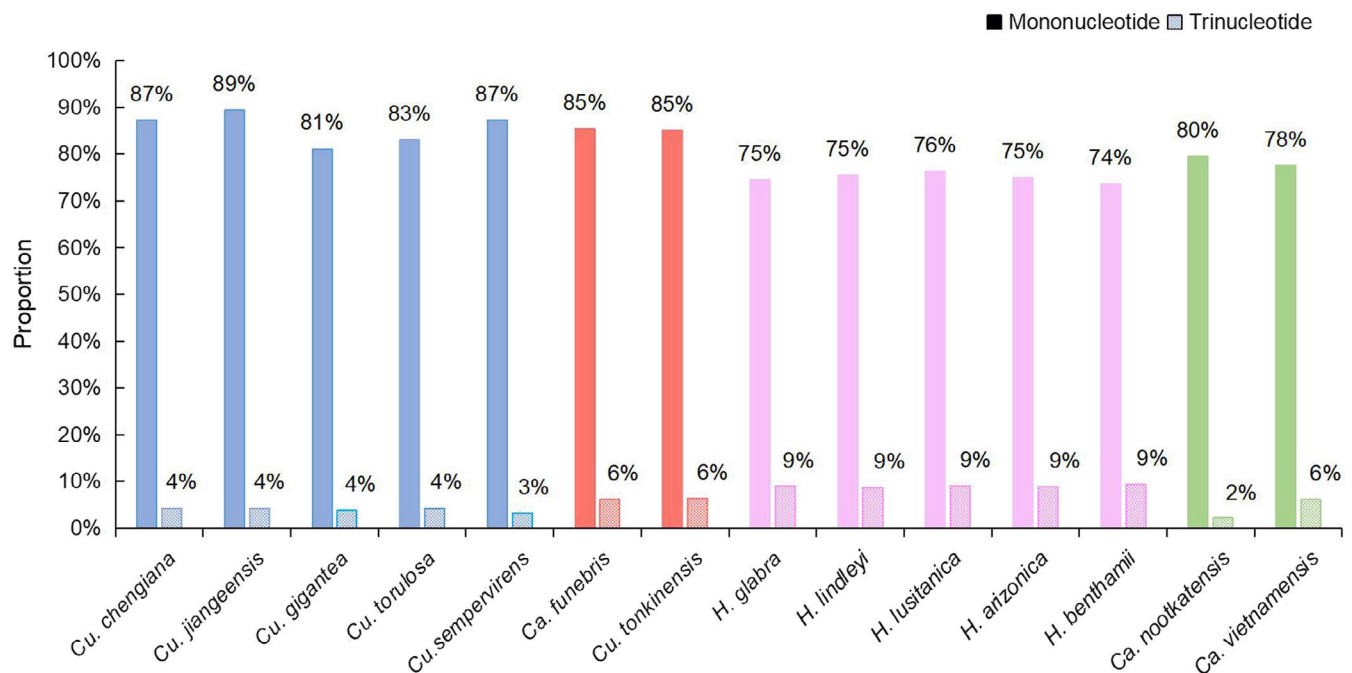
The number of repetitions is distributed between 5 and 47, which is divided into four intervals (Figure 4e). The number of 5–10 repetitions is the most (352, 48.8%), and the number of repetitions >20 is the least (7, 0.9%). The distribution of the number of repetitions in different groups has lineage characteristics (Figure 4d, Table 3). The number of 5–10 repetitions is the highest in *Hesperocyparis* (53.5%–56.1%), the number of 11–15 repeats is the highest in the *Cupressus* (50.8%–57.4%), and the number of 16–20 repeats is the highest in the *Callitropsis* (9%–14%).

Location of SSRs on the chloroplast genome was examined. 70.9% were located in the intergenic spacer region (IGS), and 18.9% were located in coding sequence region (CDS) (Figure 6a). The SSRs of each species were mainly distributed in the IGS region

(63.8%–76.4%) and less in IGS + CDS (Cross-structural) region (Figure 6b). Of the 29 SSRs distributed in IGS + CDS, 19 were located in *rps19*-IGS. The SSRs located in IGS-*atpE* occur only in *Cupressus* and *Ca. funebriis*, whereas the SSRs located in *petG*-IGS occur only in *Hesperocyparis*. There is at least one AAG/CTT distributed in the CDS region of each species, and this region is limited to *ycf1* or *rpoB*. Specifically, there is only one located at *ycf1* in *Ca. nootkatensis*, only one located at *rpoB* in *Ca. vietnamensis*, and there are 2–3 located at *ycf1* and *rpoB* in the remaining species (Appendix S1).

### 3.6 | Analysis on adaptive evolution of Cupressales

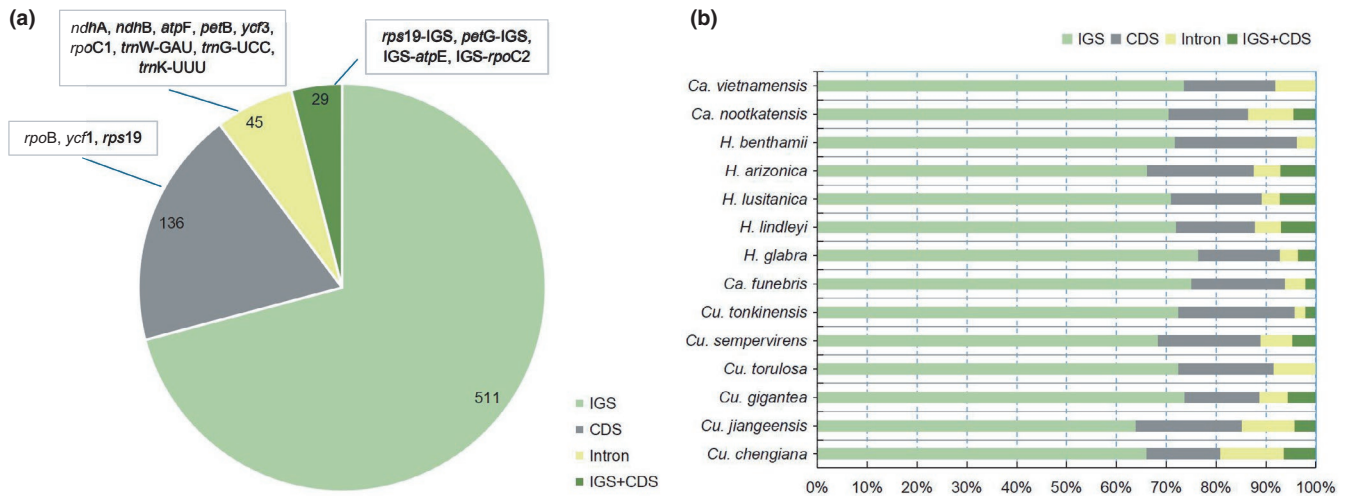
A total of 63 common genes were selected (Table 4), including 43 photosynthetic system genes, 17 genetic expression genes, and 3 miscellaneous protein genes. Based on the phylogenetic relationship of Cupressales (Figure 2) and cypresses (Figure 4a) obtained above, the phylogenetic tree of Cupressales was manually adjusted (Figure 8). Under the Branch model, the likelihood ratio test results of M0 and F showed that 60 genes accepted the M0 model ( $p > .05$ ) and 3 genes (*rps2*, *atpB*, *rpoA*) accepted the F model ( $p < .05$ ) (Figure 7). For *rps2*, *atpB* and *rpoA*, there were 15, 3, and 21 branches with the  $\omega$  value >1, respectively, and the specific branch positions were shown in Figure 8. In the likelihood ratio test of Model2 and M0 with cypresses as the foreground, there were three genes with significant



**FIGURE 5** The proportion of mononucleotide repeats and trinucleotide repeats in each species

**TABLE 3** The proportion of the number of repetitions in different groups

Repeat number	<i>Cupressus</i>	<i>Ca. funebries</i>	<i>Cupressus tonkinensis</i>	<i>Hesperocyparis</i>	<i>Ca. nootkatensis</i>	<i>Ca. vietnamensis</i>
5–10	42.6–47.2%	50%	51.1%	53.5–56.1%	43.1%	38.8%
11–15	50.8–57.4%	39.6%	36.2%	40–41.8%	47.7%	44.9%
16–20	0–4.3%	10.4%	10.6%	1.8–3.6%	9.1%	14.3%
>20	0%	0%	2%	0–4%	0%	2%

**FIGURE 6** The position distribution of the repeated sequence on the genome. (a) The distribution of all the repeating sequences. Most distributed in the intergenic spacer (IGS) region (70.9%), and at least 4% are distributed across structural regions (IGS + CDS). (b) The position distribution of the repeated sequence of each species

differences: *rps19* ( $\omega_{\text{background}} = 0.38377$ ,  $\omega_{\text{foreground}} = 0.0001$ ), *rpl20* ( $\omega_{\text{background}} = 0.38526$ ,  $\omega_{\text{foreground}} = 0.0001$ ), and *rpoC2* ( $\omega_{\text{background}} = 0.46030$ ,  $\omega_{\text{foreground}} = 0.47076$ ). The likelihood ratio test results of Model2 and M0 with the ancestral branches of *Ca. funebries* and *Cu. tonkinensis* as the foreground branch showed that only the *rpoC2* had a significant difference ( $\omega_{\text{background}} = 0.46030$  and  $\omega_{\text{foreground}} = 0.47076$ ,  $p < .05$ ). In the Site model, based on the likelihood ratio test results of M2a and M1a (Appendix S2), there were 17 genes with significant differences ( $p < .05$ ). With posterior probability  $p > 99\%$  as the standard, positive selection sites were detected in 10 genes (Table 5). The Branch-Site model detected an amino acid site under positive selection in the *rpoA* gene with *Ca. funebries* and *Cu. tonkinensis* as foreground branches (Appendix S3).

## 4 | DISCUSSION

### 4.1 | *Callitropsis funebries* is more closely related to the *Cupressus* than *Callitropsis*

Phylogenetic relationships among *Juniperus* and cypresses of Cupressaceae are controversial. Our result showed the 8 topological structures constructed for 32 species are incongruent. Particularly,

the phylogenetic relationships from complete chloroplast genome were chaotic (Figure 2a–d). Specifically, the *Hesperocyparis* and *Chamaecyparis* did not form a monophyletic branch, respectively. The possible reason is due to the multiple rearrangements of the genome (especially inversions [Wu & Chaw, 2016]) during the evolution of Cupressales. So it is difficult to align during the alignment process, thereby losing most of the informative sites and causing phylogenetic relationship chaos. In short, genome rearrangement has an impact on sequence alignment analysis. However, we observe that the monophyletic branch of *Ca. funebries* and *Cu. tonkinensis* in the four topological structures is always sister groups with *Cupressus*.

*rbcl* and *matK* are used as DNA Barcodes (Hollingsworth et al., 2011), and the obtained phylogenetic relationship has a high degree of recognition. But it is uncertain to resolve deeper branching relationships due to too few sequence differences (Tonti-Filippini et al., 2017). As it appears in the results, there is a parallel relationship among related species (Figure 2e–h). However, the phylogenetic relationship between genera is very clear and consistent with the phylogenetic relationship constructed from nuclear genes (Lu et al., 2014) or plastid genomes (Gadek et al., 2000; Hao et al., 2016; Qu, Jin, et al., 2017; Qu, Wu, et al., 2017). Also, all trees show that the monophyletic branch of *Ca. funebries* and *Cu. tonkinensis* is always sister groups with *Cupressus*. It should be noted that, in addition to the

BI tree, the other three trees all show that cypresses is a monophyletic group, but the support is not high (NJ:67, MP:48, ML:59). This also shows the complexity of cypresses and *Juniperus* (Zhu et al., 2018).

**TABLE 4** Common genes of sample species for molecular evolutionary analysis

Gene type		Gene name
Genes for photosynthesis	Photosystem I	<i>psaA psaB psaC psaI psaJ psaM</i>
	Photosystem II	<i>psbA psbB psbC psbD psbE psbF psbH psbI psbJ psbK</i>
		<i>psbL psbM psbN psbT psbZ</i>
	Cytochrome	<i>petA petB petD petG petL petN</i>
	ATP synthase	<i>atpB atpF atpI</i>
	RubiscoCO large subunit	<i>rbcL</i>
	NADH dehydrogenase	<i>ndhB ndhC ndhE ndhF ndhG ndhH ndhI ndhJ ndhK</i>
	Chlorophyll biosynthesis	<i>chlB chlL chlN</i>
Genetic system genes	Ribosomal proteins (LSU)	<i>rpl14 rpl20 rpl23 rpl33 rpl36</i>
	Ribosomal proteins (SSU)	<i>rps2 rps4 rps7 rps8 rps11 rps14 rps15 rps19</i>
	RNA polymerase	<i>rpoA rpoB rpoC1 rpoC2</i>
Others genes	Maturase	<i>matK</i>
	Envelop membrane protein	<i>cemA</i>
	c-type cytochrome synthesis	<i>ccsA</i>

## 4.2 | The complexity of cypresses (especially *Callitropsis*)

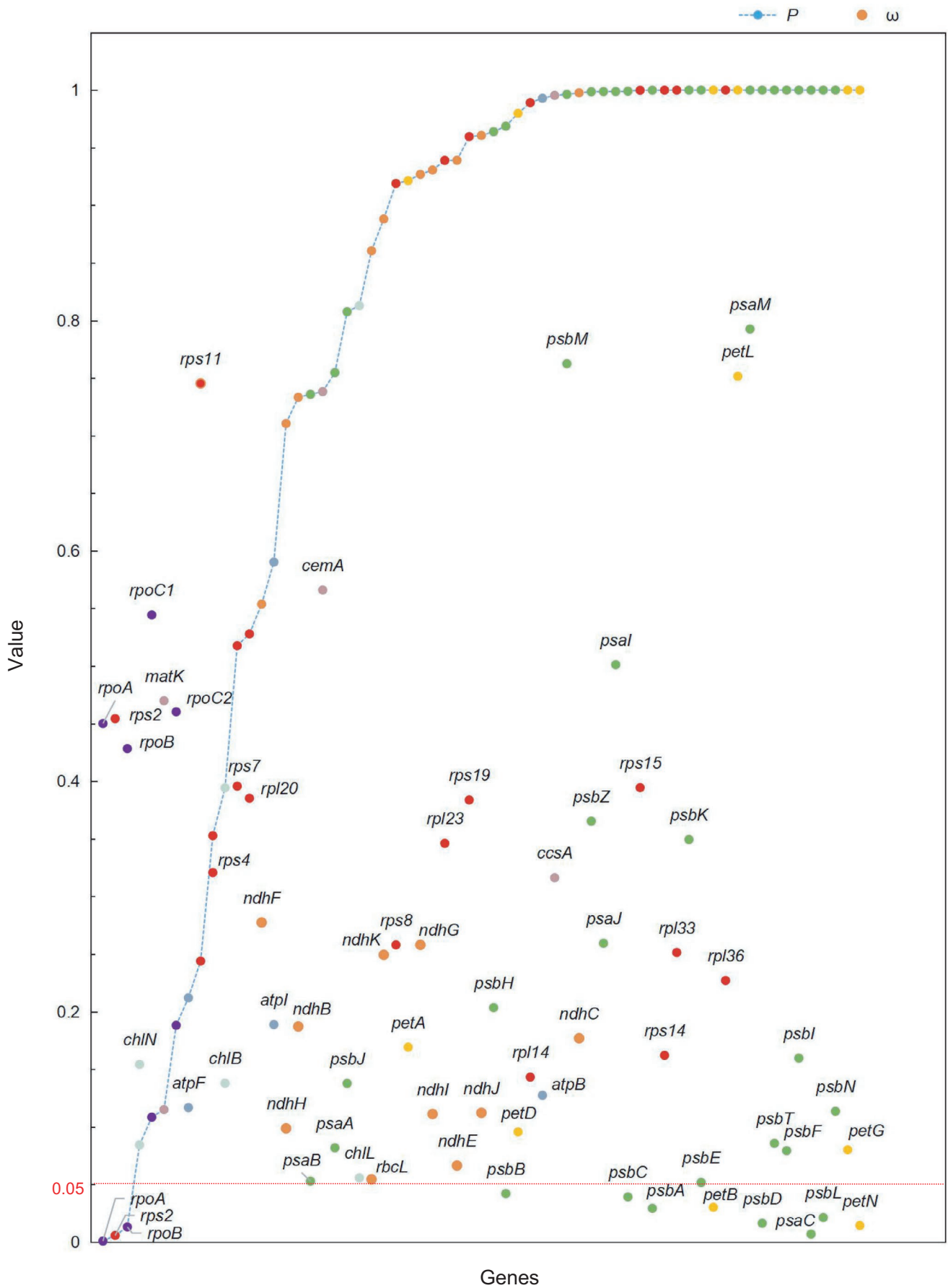
In this research, in order to further determine the phylogenetic relationship of the cypresses, the complete sequence of 15 chloroplast genomes with the same gene order was selected. All trees highly support *Hesperocyparis* and *Cupressus* to form monophyletic branches, respectively (Figure 4a), which is consistent with the results of Mao et al. (2010) based on nuclear genome sequence and chloroplast genome sequence. Consistent with Figure 2, the monophyletic branch formed by *Ca. funebris* and *Cu. tonkinensis* presents as the sister group of *Cupressus*.

Among them, the most controversial is *Callitropsis*. Farjon et al. (2002) classified a new cypress species from Vietnam into a new genus—*Xanthocyparis*, and a particular cypress species (*Cupressus nootkatensis*) into the new genus. Little (2006) speculated that the *Cupressus* is paraphyletic, and the New World cypresses diverged from a clade that later produced *Juniperus*. He assigned the New World cypresses and *Xanthocyparis nootkatensis* and *Xanthocyparis vietnamensis* to the genus *Callitropsis*. In subsequent studies, the phylogenetic relationships constructed from nuclear genome and chloroplast genome data showed that *Ca. nootkatensis* and *Ca. vietnamensis* are nested in turn at the base of the monophyletic group *Hesperocyparis* (Mao et al., 2010; Terry et al., 2012). This is consistent with the results of this study, indicating that *Callitropsis* under the existing classification system is not a monophyletic group. Because of this, recent research supports the division of 4 genera in cypresses, namely *Callitropsis*, *Cupressus*, *Hesperocyparis*, and *Xanthocyparis* (Zhu et al., 2018).

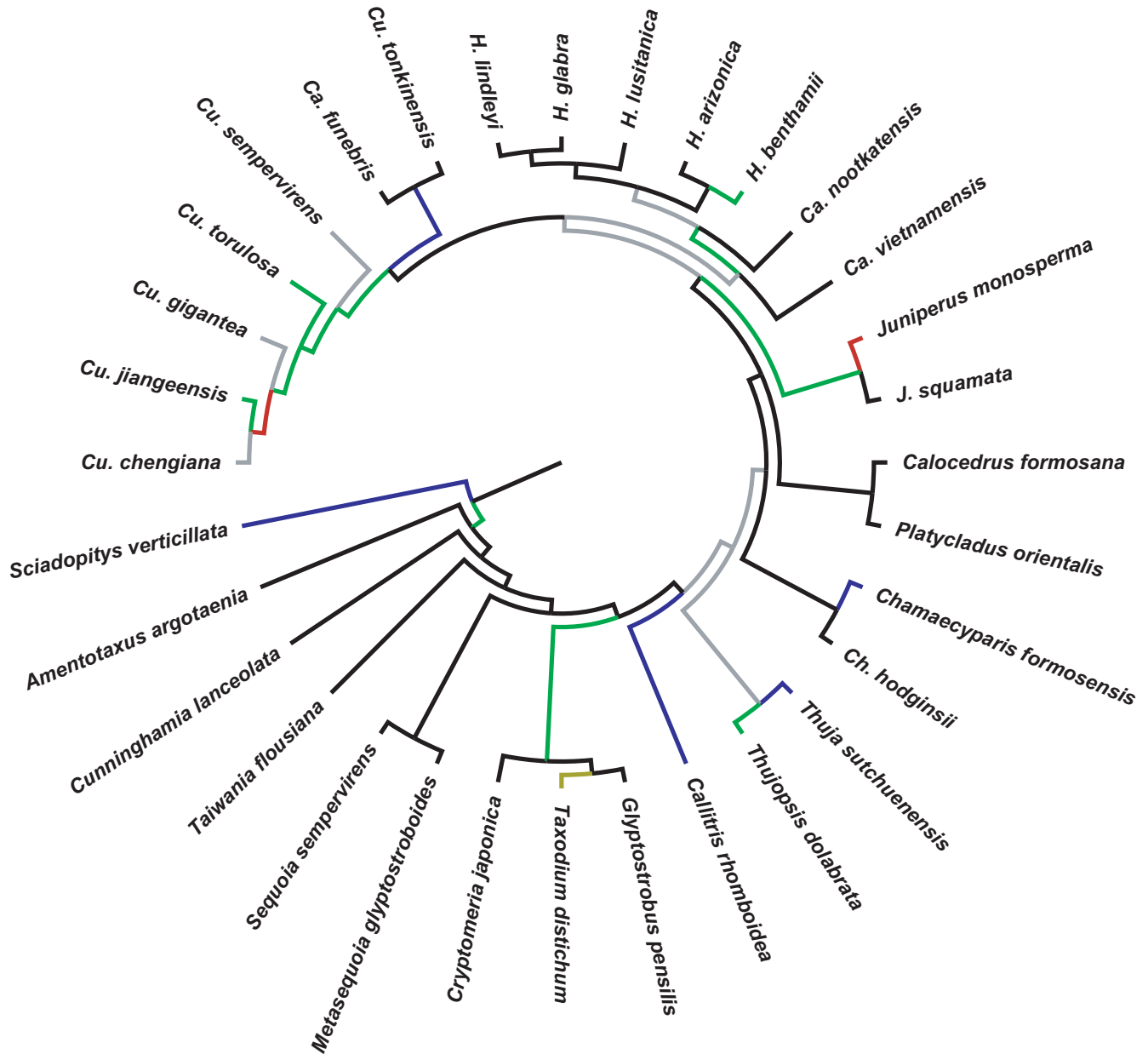
A common factor that interferes with phylogenetic analysis is hybridization events, especially when using cytoplasmic loci. If recombination occurs between different plastid haplotypes, the impact will be more serious (Wolfe & Randle, 2004). In conifers, hybridization may lead to chloroplast capture, nuclear introgression, and phylogenetic inconsistencies between the nuclear and plastid genomes (Peng & Wang, 2008; Sullivan et al., 2017; Xiang et al., 2015). Studies have found that uniparental inheritance of mitochondrial and chloroplast genomes (maternal-to-paternal and vice versa) is often reversed

Gene	Positively selected amino acid sites
<i>rpoA</i>	165S, 252S, 331G
<i>rpoB</i>	51T, 94C, 99Q, 214Y, 217K, 248E, 371L, 376T
<i>rpoC1</i>	41Q, 75P, 76M, 84Y, 166E, 178D, 234L, 240T, 241Q, 243L, 244N, 245Q, 246D, 247S, 248P, 254D, 256I, 331N, 338W
<i>rpoC2</i>	9P, 464P, 469E, 644L, 1126P, 1135I, 1147L, 1150Q, 1156L, 1159L, 1160V, 1162I, 1167K, 1168R
<i>rps7</i>	67E, 93E
<i>rps11</i>	109K
<i>rpl20</i>	116L
<i>psbM</i>	30K
<i>cemA</i>	90K, 129K
<i>ccsA</i>	205R

**TABLE 5** The positive selective amino acid sites were selected from the positive selective genes. (Take the amino acid sequence of *Ca. nootkatensis* as the reference sequence)



**FIGURE 7** Under the Branch model, the likelihood ratio of M0 and F model was tested to get the  $p$  value and  $\omega$  in M0. Three genes (*rps2*, *atpB*, and *rpoA*) accepted the F model ( $p < 0.05$ )



**FIGURE 8** Phylogenetic relationship of Cupressales. Its also a tree file in the PAML analysis. Branches with different colors represent branches with  $\omega > 1$  in the three genes in the Branch model that accept the F model. Blue: *rps2*, Red: *atpB*, Green: *rpoA*. Gray: *rpoA* and *rps2* shared. Yellow: three genes shared. The positive selective branches appear mostly in the middle and late stages of the evolution of Cupressaceae

(Whittle & Johnston, 2002), and genetic leakage has been observed in many Cupressaceae species and other seed plants (Wagner et al., 1991; Weihe et al., 2009). Therefore, cypresses may have undergone some degree of network evolution. In general, the research has enriched the plastid genomics research of cypresses, and we need to combine more effective data sets in future exploration.

### 4.3 | The unique chloroplast genomic structure of *Ca. funebris*

During the evolution of plants, the chloroplast genome often undergoes some changes, including rearrangement (Guisinger et al.,

2010; Wolf et al., 2011; Wu et al., 2007) and gene loss (Braukmann et al., 2009; Delannoy et al., 2011). These changes occur independently and are often unique to specific taxa (Wolf et al., 2010). During the evolution from *Chamaecyparis* to cypresses, the chloroplast genome structure has at least 3 insertions and inversions (Figure 3). The structural characteristics of cypresses' chloroplast genome show high consistency. Only one gene between *Cupressus* and *Hesperocyparis* is inverted. *Ca. nootkatensis* and *Ca. vietnamensis* are the same as *Cupressus* and *Hesperocyparis*, respectively, showing the particularity of *Callitropsis*. In particular, compared with *Cupressus* and *Hesperocyparis*, several genes of *Ca. funebris* were inverted, showing the uniqueness of *Ca. funebris*. Existing studies have shown that in Podocarpaceae and Taxaceae, the

number of inversions between each genera differs by 6 times (Wu & Chaw, 2016). Similarly, *Ca. nootkatensis* and *Ca. vietnamensis* should belong to different genera according to the inversion. This seems to be consistent with the views of previous studies (Zhu et al., 2018). And *Ca. funebris* should be considered as an independent genus. Moreover, from the analysis of chloroplast genome (GC content, gene number, and genome structure), the division of *Ca. funebris* and *Chamaecyparis* is not disputed.

#### 4.4 | | The distribution pattern of SSRs is beneficial to the phylogenetic relationship study

SSRs are dominated by A/T bases, and very few G/C were found and are mainly located in the IGS region. This is consistent with the existing chloroplast SSR report (Gui et al., 2020; Li et al., 2017). They show lineage specificity on the type and proportion of repeat motif and the distribution of repeat times. *Cupressus* and *Hesperocyparis* show highly uniform characteristics, respectively, while *Callitropsis* differs from them. There are differences between *Ca. funebris* and other two species in *Callitropsis*. In the proportion of A/T motif and mononucleotide repeats, *Ca. funebris* is closer to the characteristics of *Cupressus* (Figure 4b). In the SSR multidirectional comparison, *Cupressus* and *Callitropsis* have notable differences. Overall, SSRs are abundant in cypresses and can be used for intergenus identification, as well as to detect genetic diversity at the population and interspecies level.

#### 4.5 | | Adaptation changes occurred during the evolution of Cupressales

During the evolution of Cupressales, positive selection branches were detected in three genes, and the positive selection effect of these three genes occurred in the middle and early stages of Cupressales evolution. There are two genes with an obvious difference in selection pressure detected between cypresses and other branches, indicating that the two genes are experienced to strong negative selection effect in the cypresses branch.

To better understand the evolutionary history of Cupressales, the analysis of its genetic diversity and adaptive evolution is essential. Positively selected genes play an important role in the adaptation to various environments. Ten genes with positively selected sites were detected in the Site model, seven genes belonged to genetic system gene (*rpo-*, *rps7*, *rps11*, *rpl20*), 1 photosystem gene (*psbM*), and two other genes (*cemA*, *ccsA*). *Rpo-* gene encodes RNA polymerase, the most critical enzyme in transcription. *Rps7*, *rps11*, and *rpl20* encode ribosome 30S small subunit S7, S11 protein, and 50S large subunit L20 protein, respectively, all of which are involved in protein synthesis. *PsbM* is one of the components of the core complex of photosystem II. *cemA* encodes a chloroplast envelope membrane protein (Sasaki et al., 1993) and is inferred to indirectly

influence CO<sub>2</sub> uptake in plastid (Rolland et al., 1997). The *ccsA* gene encodes a protein required for heme attachment to c-type cytochromes (Merchant, 1996).

Understanding the patterns of divergence and adaptation among the members of a specific phylogenetic clade can offer important clues about the forces driving its evolution (Li et al., 2019). In this study, we detected that only the *rpoA* gene had undergone positive selection in the *Ca. funebris* branch, and only one positively selected site was detected. This may be due to plants having multiple strategies to adapt to the environment, and the adaptive modification of other abiotic stress-targeted genes in the nucleus is sufficient to maintain photosynthesis homeostasis. As a result, adaptive evolution of chloroplast-encoded genes is not required (Li et al., 2019).

All these genes might play important roles when founder effects occur in populations, both changes in selection pressure and genetic drift may result in the rapid shift of these genes to a new coadapted combination. The results provide an insight into adaptive evolution of Cupressales and a basis for further clarifying the chloroplast genetic characteristics.

## 5 | | CONCLUSIONS

Morphological-based research has a limited role in diagnosing the relationship between orders or classes of any group of plants, especially for intermediate types such as cypresses. The analysis from molecular data can provide strong support and insights for its phylogenetic analysis. Different from its morphological characteristics, the analysis results based on the phylogenetic relationship and genomic characteristics (SSRs, etc.) of the complete chloroplast genome sequences support that the monophyletic branch of *Ca. funebris* and *Cu. tonkinensis* is a sister branch of *Cupressus* (or as the base group of *Cupressus*). And *Callitropsis* is not a monophyletic group, *Ca. nootkatensis* and *Ca. vietnamensis* are nested in turn at the base of the monophyletic group *Hesperocyparis*. At the same time, *Ca. funebris* and *Chamaecyparis* are two different groups. During the evolution of Cupressales, multiple genes were detected to experience positive selection, suggesting that they may have undergone adaptive changes. One problem is that the data for constructing the topological structure only come from the chloroplast genome, which is far from enough. It is necessary to combine the nuclear genome or mitochondrial genome data, which will be more conducive to understanding their evolutionary process.

#### ACKNOWLEDGMENTS

We thank Ming Zhu for her suggestions on the manuscript and the National Natural Science Foundation for its support (31670200, 31770587, 31872670, and 32071781).

#### CONFLICT OF INTERESTS

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTION

**Jingyao Ping:** Conceptualization (equal); Data curation (equal); Writing-original draft (equal). **Peipei Feng:** Data curation (equal). **Jinye Li:** Data curation (equal); Formal analysis (equal). **Rongjing Zhang:** Resources (equal). **Yingjuan Su:** Writing-review & editing (equal). **Ting Wang:** Conceptualization (equal); Funding acquisition (equal); Writing-review & editing (equal).

## ETHICS STATEMENT

The research does not involve any ethical issues.

## DATA AVAILABILITY STATEMENT

Data source is NCBI database: [https://www.ncbi.nlm.nih.gov/nucleotide/NC\\_034788,NC\\_028155,NC\\_036939,NC\\_026296,NC\\_039562,NC\\_039563,KP099642,KX832629,KX832624,NC\\_039566,MH121051,NC\\_039565,NC\\_039564,NC\\_023121,NC\\_024022,NC\\_044076,NC\\_034943,NC\\_036996,NC\\_042176,KX832628,NC\\_034940,NC\\_034941,NC\\_031354,NC\\_010548,NC\\_030372,NC\\_027423,NC\\_021441,NC\\_021437,KX832626,NC\\_027581,NC\\_029734,MT227813](https://www.ncbi.nlm.nih.gov/nucleotide/NC_034788,NC_028155,NC_036939,NC_026296,NC_039562,NC_039563,KP099642,KX832629,KX832624,NC_039566,MH121051,NC_039565,NC_039564,NC_023121,NC_024022,NC_044076,NC_034943,NC_036996,NC_042176,KX832628,NC_034940,NC_034941,NC_031354,NC_010548,NC_030372,NC_027423,NC_021441,NC_021437,KX832626,NC_027581,NC_029734,MT227813).

## ORCID

Jingyao Ping  <https://orcid.org/0000-0003-2579-2561>

Yingjuan Su  <https://orcid.org/0000-0001-6110-2869>

Ting Wang  <https://orcid.org/0000-0001-5025-4584>

## REFERENCES

- Adams, R. P., Bartel, J., & Price, R. A. (2009). A new genus, *Hesperocyparis*, for the cypresses of the western hemisphere (CUPRESSACEAE). *Phytologia*, 91(1), 160–185.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J.-H., Zhang, Z., Miller, W., Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, 25(17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Barrett, C. F., Baker, W. J., Comer, J. R., Conran, J. G., Lahmeyer, S. C., Leebens-Mack, J. H., Li, J., Lim, G. S., Mayfield-Jones, D. R., Perez, L., Medina, J., Pires, J. C., Santos, C., Wm. Stevenson, D., Zomlefer, W. B., & Davis, J. I. (2016). Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist*, 209(2), 855–870. <https://doi.org/10.1111/nph.13617>
- Beier, S., Thiel, T., Münch, T., Scholz, U., & Mascher, M. (2017). MISA-web: A web server for microsatellite prediction. *Bioinformatics*, 33(16), 2583–2585. <https://doi.org/10.1093/bioinformatics/btx198>
- Bock, R. (2017). Witnessing genome evolution: Experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annual Review of Genetics*, 51(1), 1–22. <https://doi.org/10.1146/annurev-genet-120215-035329>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Braukmann, T. W., Kuzmina, M., & Stefanovic, S. (2009). Loss of all plastid *ndh* genes in Gnetales and conifers: Extent and evolutionary significance for the seed plant phylogeny. *Current Genetics*, 55(3), 323–337. <https://doi.org/10.1007/s00294-009-0249-7>
- Cavalier-Smith, T. (2002). Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Current Biology*, 12(2), R62–R64. [https://doi.org/10.1016/S0960-9822\(01\)00675-3](https://doi.org/10.1016/S0960-9822(01)00675-3)
- Chmielewski, M., Meyza, K., Chybicki, I. J., Działuk, A., & Burczyk, J. (2015). Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in Poland. *iForest - Biogeosciences & Forestry*, 8, 765. <https://doi.org/10.3832/IFOR1597-008>
- De Laubenfels, D. J. (2009). Nomenclatural actions for the New World cypresses (Cupressaceae). *Novon*, 19(3), 300–306. <https://doi.org/10.3417/2006176>
- De Laubenfels, D. J., Husby, C., & Griffith, M. (2012). Further nomenclatural action for the Cypresses (Cupressaceae). *Novon*, 22(1), 8–15. <https://doi.org/10.2307/23259410>
- De Santana Lopes, A., Gomes Pacheco, T., Nascimento da Silva, O., Magalhães Cruz, L., Balsanelli, E., Maltempi de Souza, E., de Oliveira Pedrosa, F., & Rogalski, M. (2019). The plastomes of *Astrocaryum aculeatum* G. Mey. and *A. murumuru* Mart. show a flip-flop recombination between two short inverted repeats. *Planta*, 250, 1229–1246. <https://doi.org/10.1007/s00425-019-03217-z>
- Delannoy, E., Fujii, S., Colas des Francs-Small, C., Brundrett, M., & Small, I. (2011). Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution*, 28(7), 2077–2086. <https://doi.org/10.1093/molbev/msr028>
- Erixon, P., & Oxelman, B. (2008). Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS One*, 3(1), e1386. <https://doi.org/10.1371/journal.pone.0001386>
- Farjon, A. (2005). *A monograph of Cupressaceae and Sciadopitys*. (648). Kew: Royal Botanic Gardens. ISBN 1842460684.
- 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. <https://doi.org/10.2307/3392953>
- Gadek, P. A., Alpers, D. L., Heslewood, M. M., & Quinn, C. J. (2000). Relationships within Cupressaceae sensu lato: A combined morphological and molecular approach. *American Journal of Botany*, 87(7), 1044–1057. <https://doi.org/10.2307/2657004>
- Gadek, P. A., & Quinn, C. J. (1987). Biflavones and the affinities of *Cupressus funebris*. *Phytochemistry*, 26(9), 2551–2552. [https://doi.org/10.1016/S0031-9422\(00\)83874-1](https://doi.org/10.1016/S0031-9422(00)83874-1)
- Gomes Pacheco, T., de Santana Lopes, A., Monteiro Viana, G. D., Nascimento da Silva, O., Morais da Silva, G., do Nascimento Vieira, L., Guerra, M. P., Nodari, R. O., Maltempi de Souza, E., de Oliveira Pedrosa, F., Otoni, W. C., & Rogalski, M. (2019). Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta*, 249(2), 563–582. <https://doi.org/10.1007/s00425-018-3023-6>
- Greiner, S., Lehwark, P., & Bock, R. (2019). Organellar Genome DRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, 47(W1), W59–W64. <https://doi.org/10.1093/nar/gkz238>
- Gui, L., Jiang, S., Xie, D., Yu, L., Huang, Y., Zhang, Z., & Liu, Y. (2020). Analysis of complete chloroplast genomes of *curcuma* and the contribution to phylogeny and adaptive evolution. *Gene*, 732, 144355. <https://doi.org/10.1016/j.gene.2020.144355>
- Guisinger, M. M., Kuehl, J. V., Boore, J. L., & Jansen, R. K. (2010). Extreme reconfiguration of plastid genomes in the angiosperm family geraniaceae: rearrangements, repeats, and codon usage. *Molecular Biology and Evolution*, 28(1), 583–600. <https://doi.org/10.1093/molbev/msq229>
- Guo, W., Grewe, F., Cobo-Clark, A., Fan, W., Duan, Z., Adams, R. P., Schwarzbach, A. E., & Mower, J. P. (2014). Predominant and substoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during Cupressophyte evolution. *Genome Biology and Evolution*, 6(3), 580–590. <https://doi.org/10.1093/gbe/evu046>
- Hao, D.-C., Chen, S.-L., & Xiao, P.-G. (2010). Molecular evolution and positive Darwinian selection of the chloroplast maturase *matK*. *Journal of Plant Research*, 123(2), 241–247. <https://doi.org/10.1007/s10265-009-0261-5>

- Hao, Z.-D., Cheng, T.-L., Zheng, R.-H., Xu, H.-B., Zhou, Y.-W., Li, M.-P., Lu, F., Dong, Y., Liu, X., Chen, J., & Shi, J. (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), e0161809. <https://doi.org/10.1371/journal.pone.0161809>
- Hasegawa, M., Zhong, B.-J., & Zhong, Y. (2009). Adaptive evolution of chloroplast genomes in ancestral grasses. *Plant Signaling & Behavior*, 4(7), 623–624. <https://doi.org/10.4161/psb.4.7.8914>
- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One*, 6(5), e19254. <https://doi.org/10.1371/journal.pone.0019254>
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- Jagel, A., & Stutzel, T. H. (2001). Zur Abgrenzung von *Chamaecyparis* Spach und *Cupressus* L. (Cupressaceae) und die systematische Stellung von *Cupressus nootkatensis* D. Don. [= *Chamaecyparis Nootkatensis* (D. DON) SPACH]. *Feddes Repertorium*, 112(3–4), 179–229. <https://doi.org/10.1002/fedr.4921120304>
- Jiménez, A. (2010). Microsatellites: A powerful genetic marker for fern research. A. Kumar, & A. Revilla-Maria (eds.), *Working with ferns: issues and applications* (pp. 207–220). New York, NY: Springer. [https://doi.org/10.1007/978-1-4419-7162-3\\_15](https://doi.org/10.1007/978-1-4419-7162-3_15)
- Kapralov, M. V., & Filatov, D. A. (2007). Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evolutionary Biology*, 7(1), 73. <https://doi.org/10.1186/1471-2148-7-73>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kwon, E. C., Kim, J. H., & Kim, N. S. (2020). Comprehensive genomic analyses with 115 plastomes from algae to seed plants: Structure, gene contents, GC contents, and introns. *Genes & Genomics*, 42(6), 553–570. <https://doi.org/10.1007/s13258-020-00923-x>
- Li, B., Lin, F.-O., Huang, P., Guo, W.-Y., & Zheng, Y.-Q. (2017). Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. *Scientific Reports*, 7(1), 10073. <https://doi.org/10.1038/s41598-017-10409-8>
- Li, J., Gao, L., Chen, S.-S., Tao, K., Su, Y.-J., & Wang, T. (2016). Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of *Sciadopityaceae*. *Scientific Reports*, 6(1), 20934. <https://doi.org/10.1038/srep20934>
- Li, Z.-H., Ma, X., Wang, D.-Y., Li, Y.-X., Wang, C.-W., & Jin, X.-H. (2019). Evolution of plastid genomes of *Holcoglossum* (Orchidaceae) with recent radiation. *BMC Evolutionary Biology*, 19(1), 63. <https://doi.org/10.1186/s12862-019-1384-5>
- Little, D. P. (2006). Evolution and circumscription of the true cypresses (Cupressaceae: *Cupressus*). *Systematic Botany*, 31(3), 461–480. <https://doi.org/10.1600/036364406778388638>
- Lu, Y., Ran, J.-H., Guo, D.-M., Yang, Z.-Y., & Wang, X.-Q. (2014). Phylogeny and divergence times of gymnosperms inferred from single-copy nuclear genes. *PLoS One*, 9(9), e107679. <https://doi.org/10.1371/journal.pone.0107679>
- Mao, K.-S., Hao, G., Liu, J.-Q., Adams, R. P., & Milne, R. I. (2010). Diversification and biogeography of *Juniperus* (Cupressaceae): Variable diversification rates and multiple intercontinental dispersals. *New Phytologist*, 188(1), 254–272. <https://doi.org/10.1111/j.1469-8137.2010.03351.x>
- Martin, G. E., Rousseau-Gueutin, M., Cordonnier, S., Lima, O., Michon-Coudouel, S., Naquin, D., de Carvalho, J. F., Aïnouche, M., Salmon, A., & Aïnouche, A. (2014). The first complete chloroplast genome of the Genistoid legume *Lupinus luteus*: Evidence for a novel major lineage-specific rearrangement and new insights regarding plastome evolution in the legume family. *Annals of Botany*, 113(7), 1197–1210. <https://doi.org/10.1093/aob/mcu050>
- Nei, M., & Kumar, S. (2000). *Molecular evolution and phylogenetics*. New York: Oxford University Press. <https://books.google.com/books?hl=en&lr=&id=vtWW9bmVd1IC&pgis=1>
- Peng, D., & Wang, X.-Q. (2008). Reticulate evolution in *Thuja* inferred from multiple gene sequences: Implications for the study of biogeographical disjunction between eastern Asia and North America. *Molecular Phylogenetics and Evolution*, 47(3), 1190–1202. <https://doi.org/10.1016/j.ympev.2008.02.001>
- Ping, J.-Y., Luo, X., Zhu, M., Zhang, R., Qian, C., Su, Y., & Wang, T. (2020). The complete chloroplast genome of *Araucaria cunninghamii* (Araucariaceae). *Mitochondrial DNA Part B Resources*, 5(3), 2952–2953. <https://doi.org/10.1080/23802359.2020.1787265>
- Qu, X.-J., Jin, J.-J., Chaw, S. M., Li, D.-Z., & Yi, T.-S. (2017). Multiple measures could alleviate long-branch attraction in phylogenomic reconstruction of Cupressoidae (Cupressaceae). *Scientific Reports*, 7, 41005. <https://doi.org/10.1038/srep41005>
- Qu, X.-J., Wu, C.-S., Chaw, S. M., & Yi, T.-S. (2017). Insights into the existence of isomeric plastomes in Cupressoidae (Cupressaceae). *Genome Biology and Evolution*, 9(4), 1110–1119. <https://doi.org/10.1093/gbe/evx071>
- Rolland, N., Dorne, A. J., Amoroso, G., Sültemeyer, D. F., Joyard, J., & Rochaix, J. D. (1997). Disruption of the plastid *ycf10* open reading frame affects uptake of inorganic carbon in the chloroplast of *Chlamydomonas*. *EMBO Journal*, 16(22), 6713–6726. <https://doi.org/10.1093/emboj/16.22.6713>
- Rushforth, K., Adams, R. P., Zhong, M., Ma, X.-Q., & Pandey, R. N. (2003). Variation among *Cupressus* species from the eastern hemisphere based on Random Amplified Polymorphic DNAs (RAPDs). *Biochemical Systematics and Ecology*, 31(1), 17–24. [https://doi.org/10.1016/S0305-1978\(02\)00073-X](https://doi.org/10.1016/S0305-1978(02)00073-X)
- Saina, J., Li, Z.-Z., Gichira, A., & Liao, Y.-Y. (2018). The complete chloroplast genome sequence of tree of heaven (*Ailanthus altissima* (Mill.) (Sapindales: Simaroubaceae), an important pantropical tree. *International Journal of Molecular Sciences*, 19(4), 929. <https://doi.org/10.3390/ijms19040929>
- Sasaki, Y., Sekiguchi, K., Nagano, Y., & Matsuno, R. (1993). Chloroplast envelope protein encoded by chloroplast genome. *FEBS Letters*, 316(1), 93–98. [https://doi.org/10.1016/0014-5793\(93\)81743-J](https://doi.org/10.1016/0014-5793(93)81743-J)
- Stamatakis, A. (2014). RaxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Sullivan, A. R., Schiffthaler, B., Thompson, S. L., Street, N. R., & Wang, X.-R. (2017). Interspecific plastome recombination reflects ancient reticulate evolution in *Picea* (Pinaceae). *Molecular Biology and Evolution*, 34, 1689–1701. <https://doi.org/10.1093/molbev/msx111>
- Swofford, D. L. (2002). PAUP\*: Phylogenetic analysis using parsimony (and other methods), version 4. Sunderland, MA: Sinauer Associates.
- Terry, R. G., Bartel, J. A., & Adams, R. P. (2012). Phylogenetic relationships among the New World cypresses (*Hesperocyparis*: Cupressaceae): Evidence from noncoding chloroplast DNA sequences. *Plant Systematics and Evolution*, 298(10), 1987–2000. <https://doi.org/10.1007/s00606-012-0696-3>
- Tonti-Filippini, J., Nevill, P. G., Dixon, K., & Small, I. (2017). What can we do with 1000 plastid genomes? *The Plant Journal*, 90(4), 808–818. <https://doi.org/10.1111/tpj.13491>



- Vieira, L. N., Rogalski, M., Faoro, H., Fraga, H. P., Anjos, K. G., Picchi, G. F. A., Nodari, R. O., de Oliveira Pedrosa, F., De Souza, E. M., & Guerra, M. P. (2016). The plastome sequence of the endemic Amazonian conifer, *Retrophyllum piresii* (Silba) C.N.Page, reveals diferent recombination events and plastome isoforms. *Tree Genetics and Genomes*, 12(1), 10. <https://doi.org/10.1007/s11295-016-0968-0>
- Wagner, D. B., Dong, J., Carlson, M. R., & Yanchuk, A. D. (1991). Paternal leakage of mitochondrial DNA in *Pinus*. *Theoretical and Applied Genetics*, 82(4), 510–514. <https://doi.org/10.1007/BF00588607>
- Wang, L.-S., Yu, J., Zhang, X.-C., & Qin, H.-N. (2015). Overview of higher plant diversity in China. *Biodiversity Science*, 23, 217–224. <https://doi.org/10.17520/biods.2015049>
- Wang, X.-Q., & Ran, J.-H. (2014). Evolution and biogeography of gymnosperms. *Molecular Phylogenetics and Evolution*, 75, 24–40. <https://doi.org/10.1016/j.ympev.2014.02.005>
- Weihe, A., Apitz, J., Pohlheim, F., Salinas-Hartwig, A., & Borner, T. (2009). Biparental inheritance of plastidial and mitochondrial DNA and hybrid variegation in *Pelargonium*. *Molecular Genetics and Genomics*, 282(6), 587–593. <https://doi.org/10.1007/s00438-009-0488-9>
- Whittle, C. A., & Johnston, M. O. (2002). Male-driven evolution of mitochondrial and chloroplast DNA sequences in plants. *Molecular Evolution and Evolution*, 19(6), 938–949. <https://doi.org/10.1093/oxfordjournals.molbev.a004151>
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, 76(3–5), 273–297. <https://doi.org/10.1007/s11103-011-9762-4>
- Wolf, P. G., Der, J. P., Duffy, A. M., Davidson, J. B., Grusz, A. L., & Pryer, K. M. (2011). The evolution of chloroplast genes and genomes in ferns. *Plant Molecular Biology*, 76(3–5), 251–261. <https://doi.org/10.1007/s11103-010-9706-4>
- Wolf, P. G., Roper, J. M., & Duffy, A. M. (2010). The evolution of chloroplast genome structure in ferns. *Genome*, 53(9), 731–738. <https://doi.org/10.1139/g10-061>
- Wolfe, A. D., & Randle, C. P. (2004). Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: Implications for plant molecular systematics. *Systematic Botany*, 29(4), 101–120. <https://doi.org/10.1600/0363644042451008>
- Wu, C.-S., & Chaw, S. M. (2016). Large-scale comparative analysis reveals the mechanisms driving plastomic compaction, reduction, and inversions in Conifers II (Cupressophytes). *Genome Biology and Evolution*, 8(12), 3740–3750. <https://doi.org/10.1093/gbe/evw278>
- Wu, C.-S., Wang, Y.-N., Liu, S.-M., & Chaw, S. M. (2007). Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: Insights into cpDNA evolution and phylogeny of extant seed plants. *Molecular Biology and Evolution*, 24(6), 1366–1379. <https://doi.org/10.1093/molbev/msm059>
- Wyman, S. K., Jansen, R. K., & Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, 20(17), 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Xia, X. (2017). DAMBE6: New tools for microbial genomics, phylogenetics and molecular evolution. *Journal of Heredity*, 108(1), 431–437. <https://doi.org/10.1093/jhered/esx033>
- Xiang, Q.-P., Wei, R., Shao, Y.-Z., Yang, Z.-Y., Wang, X.-Q., & Zhang, X.-C. (2015). Phylogenetic relationships, possible ancient hybridization, and biogeographic history of *Abies* (Pinaceae) based on data from nuclear, plastid, and mitochondrial genomes. *Molecular Phylogenetics and Evolution*, 82(82), 1–14. <https://doi.org/10.1016/j.ympev.2014.10.008>
- Xie, Z., & Merchant, S. (1996). The plastid-encoded *ccsA* gene is required for heme attachment to chloroplast c-type cytochromes. *Journal of Biological Chemistry*, 271(9), 4632–4639. <https://doi.org/10.1074/jbc.271.9.4632>
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z., & dos Reis, M. (2010). Statistical properties of the branch-site test of positive selection. *Molecular Biology and Evolution*, 28(3), 1217–1228. <https://doi.org/10.1093/molbev/msq303>
- Yang, Z., Swanson, W. J., & Vacquier, V. D. (2000). Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution*, 17(10), 1446–1455. <https://doi.org/10.1093/oxfordjournals.molbev.a026245>
- Yang, Z., Wong, W.-S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4), 1107–1118. <https://doi.org/10.1093/molbev/msi097>
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5), 568–573. <https://doi.org/10.1093/oxfordjournals.molbev.a025957>
- Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, 18(5), 821–829. <https://doi.org/10.1101/gr.074492.107>
- Zhang, Z.-H., An, M.-L., Miao, J.-L., Gu, Z.-Q., Liu, C., & Zhong, B.-J. (2018). The Antarctic sea ice alga *Chlamydomonas* sp. ICE-L provides insights into adaptive patterns of chloroplast evolution. *BMC Plant Biology*, 18(1), 53. <https://doi.org/10.1186/s12870-018-1273-x>
- Zheng, W.-J., & Fu, L.-G. (1978). Cupressaceae. In Wu, Z.-Y. (Ed.), *Flora of China* (Vol. 7). Beijing, BJ: Science Press. (in Chinses).
- Zhu, A.-D., Fan, W.-S., Adams, R. P., & Mower, J. P. (2018). Phylogenomic evidence for ancient recombination between plastid genomes of the *Cupressus* - *Juniperus* - *Xanthocyparis* complex (Cupressaceae). *BMC Evolutionary Biology*, 18, 137. <https://doi.org/10.1186/s12862-018-1258-2>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Ping J, Feng P, Li J, Zhang R, Su Y, Wang T. Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. *Ecol Evol*. 2021;11:4786–4802. <https://doi.org/10.1002/ece3.7381>