Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complete chloroplast genome sequence of Cupressus funebris and its comparative analysis among nine other Cupressaceae species

Sen Cao
sailingcs@163.com
Chinese Academy of Forestry

Zhen Zhang
zhenzh19860516@163.com
Chinese Academy of Forestry

Zhichun Zhou
zczhou_risf@163.com
Chinese Academy of Forestry

Haizhen Zhang
haizhenzhang@hotmail.com
Hangzhou Botanical Garden

Xiuli Chu
xiulic0207@163.com
Shanghai Botanical Garden

Ting Wang
164314353@qq.com
Hangzhou Botanical Garden

Junzhe Gu
gujunzhe@shbg.org
Shanghai Botanical Garden

Research Article

Keywords:

DOI: https://doi.org/

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Complete chloroplast genome sequence of *Cupressus funebris* and its comparative analysis among nine other Cupressaceae species

Sen Cao[a,b,c], Zhen Zhang[a,b,c,*], Zhichun Zhou[a,b,c], Haizhen Zhang[d], Xiuli Chu[e], Ting Wang[d], Junzhe Gu[e]

[a] *Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China*

[b] *State Key Laboratory of Tree Genetics and Breeding, Chinese Academy of Forestry, Beijing 100091, China*

[c] *Zhejiang Provincial Key Laboratory of Tree Breeding, Hangzhou 311400, China*

[d] *Hangzhou Botanical Garden, Hangzhou 310007, China*

[e] *Shanghai Botanical Garden, Shanghai 200000, China*

[*] *Corresponding author.* E-mail: zhenzh19860516@163.com

ABSTRACT

Cupressus funebris Endl., a widely cultivated evergreen species in China, plays an important ecological and economic role in afforestation, ecological restoration, and the production of essential oils and medicinal compounds. However, its taxonomic status within Cupressaceae has long been debated due to complex phylogenetic relationships. In this study, we sequenced and annotated the complete chloroplast genome of *C. funebris* and conducted comparative analyses with nine other Cupressaceae species. The genome of *C. funebris* is 127,816 bp in length, encoding 117 genes. Comparative analyses revealed a highly conserved genome structure across the ten species, with notable divergence concentrated in coding regions such as *accD*, *ycf1*, *ycf2*, and *rpoC2*. Codon usage analysis indicated a significant bias toward codons ending in A or U. A total of 709 simple sequence repeats (SSRs) and 170,475 single nucleotide polymorphisms (SNPs) were identified, providing valuable markers for genetic studies. The phylogenetic tree constructed from 63 complete chloroplast genomes revealed species relationships largely consistent with traditional taxonomy, indicating that chloroplast genomes can effectively reflect phenotypic variation within Cupressaceae and thus hold broad potential for systematic and evolutionary studies. Furthermore, variable regions within the *rpoA* and *rpoC2* genes were identified as SNP-rich hotspots, enabling species-level discrimination across all ten species. These findings not only enhance our understanding of chloroplast genome evolution

in Cupressaceae but also offer genomic resources for taxonomy, phylogenetics, and molecular breeding.

Keywords: *Cupressus funebris*, chloroplast, genome, Cupressaceae, SNP

Hightlight:

1、Amino acids involved in protein synthesis exhibited a pronounced codon usage bias toward codons ending in “A” or “U.”

2、A total of 709 Simple sequence repeats and 170,475 Single nucleotide polymorphisms were developed across the 10 Cupressaceae species.

3、The phylogenetic clustering based on chloroplast genomes was largely consistent with traditional taxonomy.

4、Multiple loci within the *rpoA* and *rpoC2* genes were identified that, in combination, enable the differentiation of all 10 Cupressaceae species.

1. Introduction

As one of the most essential organelles, the chloroplast possesses an independent genome that encodes chloroplast-specific components [1]. In contrast to those encoded by nuclear DNA, a small proportion of chloroplast proteins are encoded by the chloroplast genome and synthesized through its own transcription and translation processes [2]. Comparative analyses of chloroplast genome structure have revealed that most chloroplast genomes are highly conserved in both gene order and gene function [3, 4], generally maintain a relatively stable genomic architecture [5, 6]. These stable genetic feature provide valuable resources for studies in phylogenetics, DNA barcoding, taxonomy and species evolution [7-9]. Currently, chloroplast genomic data have increasingly served as valuable resources for assessing plant genetic variation, advancing conservation genetics, and facilitating the development of molecular markers [10-13].

The broad Cupressaceae family contains approximately about 162 species across 32 genera with a worldwide distribution [14-16]. However, due to similarities in phenotypic traits, coupled with relictual characteristics in biogeography and paleobotany [17, 18], the classification system of the broadly defined Cupressaceae has long been subject to debate. The current broad circumscription of the Cupressaceae was revised by Gadek et al., based on a phylogenetic tree generated using two chloroplast DNA regions (*matK* and *rbcL*) [19]. In 2018, Li et al., (2002) analyzed the evolutionary rate stability of *matK* and *rbcL* using a relative rate test, obtaining results consistent with those of Gadek et al [20]. Between these two studies, the broad concept of Cupressaceae underwent several revisions through the work of Little and other researchers [21-23]. These successive revisions highlight the taxonomic

complexity of the family.

Cupressus funebris Endl., the *Cupressus* L. genus in the family Cupressaceae, comprises evergreen tree species with a long history of cultivation in China. It was primarily distributed in southern China and was known for its strong ecological adaptability [24, 25]. As a key afforestation species on barren and degraded mountainous lands, *C. funebris* plays an important role in ecological restoration and environmental protection. In addition, *C. funebris* is also regarded as a resource-oriented species due to its production of cypress essential oils, fragrances, and medicinal products [26]. However, the taxonomic status and species diversity of the genus *Cupressus* L. had long been focal points of debate in the study of conifer systematics and evolution [27]. In recent years, with advances in molecular systematics and population genetics, both the circumscription of *Cupressus* and the number of species it comprises had remained contentious. Little (2006) and Adams et al. (2009) reclassified *Cupressus* L., *Juniperus* L., *Callitropsis* Oerst. and *Xanthocyparis* Farjon & T. H. Nguyen using nuclear gene sequences and chloroplast DNA sequences, respectively [21, 22]. With the advancement of high-throughput sequencing technologies, chloroplast genome and transcriptome sequencing have provided new insights into the phylogeny of *Cupressus* L [27-29], providing greater possibilities for the identification and classification of *C. funebris* and related species.

In this study, the chloroplast genomes of *C. funebris* and nine other Cupressaceae species were sequenced, with the sequence and structural variations were analyzed and conducted phylogenetic analyses with the chloroplast genomes of 55 previously published Cupressaceae species to clarify the phylogenetic position of *C. funebris* within the family. Additionally, a set of highly polymorphic SSR and SNP molecular markers were developed for species identification and gerplasm screeing.

2. Materials and methods

2.1 Sample collection and DNA sequencing

Fresh leaf samples from 10 Cupressaceae species were collected from Hangzhou and Shanghai, China, and immediately preserved in liquid nitrogen. Species identification was conducted by Ting Wang from Hangzhou Botanical Garden and Junzhe Gu from Shanghai Botanical Garden (Table S1). In addition, chloroplast genome sequences of 53 Cupressaceae species were retrieved from NCBI for phylogenetic tree construction. Detailed information is provided in Table S2.

Total DNA was isolated from leaves (100 mg) using the modified CTAB method (Aidlab, China). DNA concentration and purity were assessed with a microspectrophotometer (NanoDrop-2000, Thermo, USA). All samples were sequenced at Tianjin Smart Genomics Corporation (Tianjin, China).

2.2 Chloroplast genome assembly and annotation

The Illumina platform was used for sequencing, and the chloroplast genomes were assembled from the sequencing data using the GetOrganelle software (v1.7.5) with default parameters [30], resulting in circular chloroplast genome sequences. Gene annotation of the chloroplast genomes was performed using CPGAVAS2 [31], and genome visualization was carried out using OGDRAW [32]. Annotation of tRNA genes was conducted with tRNAscan-SE [33], and rRNA genes were annotated using BLASTN [32]. The complete chloroplast genome sequences and annotation information for all ten species have been submitted to the GenBank public database and the accession number are shown in Table S3.

2.3 Comparative chloroplast genome and synteny analysis

Suquence similarity analysis and visualization of the chloroplast genomes of 10 Cupressaceae species were performed using the mVISTA program (https://genome.lbl.gov/vista/mvista/ submit.shtml) under the ShuffleLaGAN mode [34]. Synteny analysis of chloroplast genome sequences and genes among 10 Cupressaceae species was conducted using the Mauve software [35], and the results were visualized accordingly. Sequence alignments of the *rpoA* and *rpoC2* genes were performed using the MUSCLE algorithm implemented in SnapGene (Ver. 6.0.2) [36].

2.4 Repeat sequences and SSR analysis

Tandem, palindromic, forward, reverse and complementary repeats were identified in the ten Cupressaceae species chloroplast genomes by the REPuter (https://bibiserv.cebitec.uni-bielefeld.de/reputer) [37]. The SSR loci were identified by MISA (https://webblast.ipk-gatersleben.de/misa/) [38]. The minimum numbers were: ten repeats for monomer; five repeats for dimer; four repeats for trimer and three repeats for tetramer, pentamer and hexamer. The shortest size between any two SSR loci was more than 150 bp.

2.5 Phylogenetic relationship analysis

Multiple sequence alignments were conducted using MAFFT software [39], followed by phylogenetic tree reconstruction using IQ-TREE [40]. The resulting phylogenetic relationships were visualized using iTOL platform (https://itol.embl.de/) [41].

3. Results

3.1. Features of chloroplast organization and structure

The size of ten complete chloroplast genomes of Cupressaceae ranged from 127,074bp (*Platycladus orientalis 'Sieboldii'*) to 129,175bp (*Thujopsis dolabrata*)(Fig. 1, Table 1 and Table 2). Among the ten Cupressaceae species, all ten chloroplast genomes were extremely conserved in GC contents and rRNA gene number, which ranged from 34.67% ~ 35.01% and presented 4 rRNA-coding genes called *rrn*4.5S, *rrn*5S, *rrn*16S and *rrn*23S. Besides, six species including *C. funebris* were extremely conserved in gene number, which presented 117 genes, including protein-coding gene, rRNA gene and tRNA gene. Particularly, *C. funebris* and *C. sempervirens* both have 82 protein-coding gene while *Chamaecyparis pisifera 'Squarrosa'*, *Chamaecyparis pisifera 'Filifera'*, *Chamaecyparis obtusa* and *Thujopsis dolabrata* have an additional gene called *rps16*. In the tRNA-coding gene set, compared to *C. funebris*, *'Squarrosa'*, *'Filifera'* and *Cha. obtuse* lack *trn*T-GGU, *T. dolabrata* lack *trn*P-GGG, *Juniperus formosana* lack *trn*P-GGG and *trn*L-CAA, while *Hesperocyparis arizonica and Hesperocyparis lusitanica lack trnf*M-CAU.

3.2. Comparative structural analysis of ten Cupressaceae chloroplast genomes

The chloroplast genomes of 10 Cupressaceae were compared by using the MAUVE program (Fig. 2). The collinearity analysis divided the 10 species into three groups. The structure of chloroplast genomes in *C. funebris*, *C. sempervirens*, *H. arizonica* and *H. lusitanica* are extremely similar and have good collinearity. *J. formosana* and *P. orientalis 'Sieboldii'* clustered together into another group, while the remaining species, including *C. pisifera 'Squarrosa'*, *C. pisifera 'Filifera'*, *C. obtuse* and *T. dolabrata*, also displayed good collinearity within their group.

In order to evaluate the sequence divergence among the ten Cupressaceae chloroplast genomes, A sequence similarity comparison using mVISTA (Frazer et al., 2004) was conducted with the plastome of *C. funebris* designated as the reference genome in Fig. 3. The results showed varying degrees of sequence similarity among different species, with most of the sequence divergence was mainly concentrated in the protein-coding regions, particularly in the exonic regions of genes(exon) such as *acc*D, *ycf1*, *ycf2* and *rpo*C2 (Fig. 3). Overall, the analyzed sequences exhibited a relatively high level of sequence divergence across the Cupressaceae family.

3.3. Codon usage analysis

The relative synonymous codon usage (RSCU) values were estimated for 10 Cupressaceae species (Fig. 4). The RSCU value of this codon is equal to one if codon use is not preferred, whereas a value greater than 1 suggests that the codon is used more frequently than expected, and vice versa. In these chloroplast genomes, A total of 64 codons-including the stop codons UAA, UGA and UAG-were identified, encoding 21 amino acids involved in protein biosynthesis. The corresponding RSCU values for each codon are presented in

Table S4. Statistical analysis of the codon usage revealed that among all codons, UUA (encoding leucine, Leu) was the most frequently used, with an average RSCU value of 2.051, whereas CUG (encoding Leu) was the least used, with an average RSCU value of only 0.315.

Overall, 32 codons and one stop codon (UAA) exhibited RSCU values greater than 1, among which 30 codons ended with either A or U, accounting for 93.75% of the total (Table S4). However, the codon AUA (encoding isoleucine, Ile) showed an RSCU value less than 1 only in *C. obtusa*, resulting in only 29 codons ending with A or U. These findings indicate that the chloroplast genomes of Cupressaceae species exhibit a clear codon usage bias, with a preference for codons ending in A or U.

3.4 Long repeats and simple sequence repeat (SSR) analysis

Five types of long repeats-tandem, palindromic, forward, reverse and complementary-were illustrated in Fig. 5. The *C. obtuse* possessed the highest number of long repeats, totaling 775, whereas *P. orientalis 'Sieboldii'* had the fewest, with only 247 repeats identified. Analysis of the variation in long repeats revealed that forward repeats accounted for the highest proportion at 53.48%, while complementary repeats exhibited the lowest proportion at 5.50% (Table S5).

Simple Sequence Repeats (SSR) analysis revealed the presence of six major types of SSRs across the ten Cupressaceae species (Table 3). A total of 709 SSRs were identified, among which *C. pisifera 'Filifera'* exhibited the highest number, with 86 markers developed. The motif compositions of SSRs varied among species. Monomer repeats were the most abundant in all ten species, accounting for 36.51% to 62.67% of the total SSRs, followed by dimer repeat (11.48% ~ 38.10%), trimer repeat (5.81% ~ 19.23%) and tetramer repeat (9.26% ~ 16.18%). Moreover, Pentamer and Hexamer repeats were concurrently identified in only three species-*C. pisifera 'Squarrosa'*, *C. pisifera 'Filifera'* and *J. formosana*-constituting merely 2.12% of all SSRs. Furthermore, a total of 170,475 SNP markers were developed from the chloroplast genomes using different species as references, providing a valuable resource for future studied such as genome-wide association studies (GWAS).

3.5 Phylogenetic analysis

A total of 63 Cupressaceae species chloroplast genomes were constructed in Fig. 6. The overall clustering results based on the chloroplast genome were largely consistent with traditional taxonomy. However, *C. sempervirens* from the genus *Cupressus* and *Juniperus chinensis* from the genus *Juniperus* were not clustered within their respective genera. 22 species in the genus Juniperus were clustered together, representing the largest cluster. *C. funebris* showed close genetic affinity to other *Cupressus* species but did not cluster within the same branch. These findings suggested that chloroplast genomes could reflect interspecific phylogenetic relationships to a certain extent and might serve as a preliminary substitute for nuclear genome data in evolutionary studies.

3.6 SNP development and species identification based on the

chloroplast genomes

This study developed species-specific molecular markers based on the highly variable regions of the *rpoA* and *rpoC2* genes of the chloroplast genomes of 10 Cupressaceae species, enabling precise taxonomic identification. In *rpoA*, the sequences of the 10 species exhibited 10 SNP sites in 139 bp, which could be used as molecular markers to distinguish any pair of species (Fig. 7a, a~j). Marker **a**, **b**, **c** and **g** were capable of differentiating the four species: *C. pisifera 'Squarrosa'*, *C. pisifera 'Filifera', H. arizonica* and *H. lusitanica*, marker **e** and **j** could distinguish between *C. sempervirens* and *J. formosana*. While each of the marker **d**, **f**, **h** and **i** was uniquely present in *C. obtusa*, *C. funebris*, *T. dolabrata* and *P. orientalis 'Sieboldii'*, respectively. In *rpoC2*, ten species could be distinguished by 8 SNPs located within a 106 bp region (Fig. 7b, a~h). Marker **a**, **b** and **g** could be jointly analyzed to distinguish among *T. dolabrata*, *H. arizonica* and *H. lusitanica*, while marker **c** and **h** could distinguish among *C. pisifera 'Squarrosa'*, *C. pisifera 'Filifera'* and *P. orientalis 'Sieboldii'*. Furthermore, marker **d**, **e** and **f** was uniquely present in *J.formosana*, *C. funebris* and *C. obtusea*, respectively.

4. DISCUSSION

4.1. Chloroplast genome structure

Advancements in next-generation sequencing (NGS) technologies have enabled the comprehensive assembly of complete chloroplast genomes and facilitated the development of molecular markersa. In this study, the chloroplast genomes of 10 Cupressaceae species were sequenced and comparatively analyzed for structural variation, and a total of 82 shared protein-coding genes were identified. Among these genes, the functions of most protein-coding genes had been well characterized; however, genes with currently unknown functions-such as *ycf1*, *ycf2*, and *ycf4*-were also widely present in the chloroplast genomes of various Cupressaceae species. Sequence divergence analyses revealed that sequence variations among these species were primarily associated with these genes. Although widely detected across most gymnosperms, the biological functions of these genes remain unclear.

4.2. Codon usage analysis

The codon usage frequency and bias in the chloroplast genomes of 10 Cupressaceae species were systematically analyzed using relative synonymous codon usage (RSCU). The results revealed notable codon usage differences both among genes within the same species and across different species, albeit driven by distinct factors. Intraspecific variation in codon preference among genes has been demonstrated to correlate with gene expression levels, amino acid composition, and protein level noise considerations [42]. By contrast, interspecific codon usage differences observed in this study are likely shaped by broader influences, including overall genomic GC content [43] and environmental factors such as optimal growth temperature [44]. Notably, codon usage convergence was apparent

among phylogenetically related taxa or those sharing similar tRNA repertoires [42], which might underlie the pronounced bias toward codons ending in A or U consistently observed across all 10 Cupressaceae species.

4.3. SSR markers

SSRs are considered ideal molecular markers due to their widespread distribution across the genome, high interspecific polymorphism, locus-specific co-dominance, and broad transferability [45]. Recently, the acquisition of chloroplast genome data through next-generation sequencing (NGS) technologies has provided a robust foundation for the high-throughput screening of SSR loci, thereby facilitating the efficient development of SSR markers across a wide range of plant taxa [46, 47]. In this study, a total of 709 SSRs were identified across the chloroplast genomes of 10 Cupressaceae species (Table 3), and these loci exhibited variations among different species. The abundant SSR loci and potential species-specific polymorphic SSRs provide valuable genetic information and sequence resources for subsequent studies on species identification, genetic diversity assessment, and phylogenetic research in Cupressaceae [48].

4.4. Phylogenetic analysis

With the advancement of NGS technologies, chloroplast genomes have been extensively utilized to resolve plant phylogenetic relationships and elucidate evolutionary patterns [49, 50]. The phylogenetic tree revealed the evolutionary positions of the 10 Cupressaceae species and their relationships with other members of the family. The results demonstrated that the genetic relationships among most species were relatively stable, thereby providing valuable insights for refining the taxonomic framework of the Cupressaceae. The number of species within the genus *Cupressus* remains unresolved. Farjon (2005) recognized a total of 15 species in *Cupressus*, while Little (2006) considered the narrowly defined *Cupressus* to include only 12 species [17, 22]. In 2018, Terry et al. (2018) reconstructed the phylogenetic relationships of Cupressus based on Little's 12 species and three species proposed by Silba (2005), resulting in a total of 15 taxa [23, 51, 52]. For the sake of taxonomic stability, the classification system of 12 species proposed by Little (2006) is still commonly adopted [22].

Except for *C. sempervirens*, the clustering results based on chloroplast genome sequences were largely congruent with traditional taxonomic classifications. Recent phylogenomic studies focusing on the narrowly defined *Cupressus* s.s., *Hesperocyparis*, *Callitropsis*, and *Xanthocyparis*, as well as broader analyses across the family Cupressaceae and even gymnosperms as a whole, consistently support a sister-group relationship between *Cupressus* s.s. and *Juniperus* [15, 27, 29, 53]. This well-supported phylogenetic pattern provides a plausible explanation for the unexpected placement of *C. sempervirens*—the type species of *Cupressus* s.s.—within the *Juniperus* clade in our phylogenetic tree. Furthermore, we re-examined the phenotype of *C. sempervirens* from the Shanghai Botanical Garden to ensure accurate species

identification (Fig. 8).

4.5. The *rpoA* and *rpoC2* as a DNA marker used for 10 Cupressaceae species identification

SNP markers are short and relatively conserved fragment sequences in the chloroplast genome sequence and distinguish species by stable mutation sites. SNP markers are now widely utilized in various fields, including the assessment of genetic diversity and structure within populations, plant species identification and classification, and studies of population genetics [54-57]. Among the coding regions, the most variable genes include *matK*, *rpoC1*, *rpoC2*, *ycf1*, *ycf2*, *ndhD* and *ndhF*, which have also been reported as highly divergent regions in previous studies [58]. These regions represent conserved segments with clustered variations rich in SNPs and indels, commonly referred to as “mutation hotspots”. To date, the *rpoC1*, *rpoC2* and *ycf1* genes have been confirmed as mutation hotspot regions in multiple species [59-61].

Based on a preliminary sequence comparison of 82 shared genes across 10 Cupressaceae species, in addition to *rpoC2*, *rpoA* was also identified as a mutation hotspot region in this study which was known to function primarily in DNA replication and Transcription [62]. In Cupressaceae, the high mutability and evidence of positive selection in *rpoA* may have contributed to adaptive evolution through plant self-replication under diverse environmental conditions [63, 64]. Further sequence alignment revealed SNP markers within *rpoA* and *rpoC2* that can be effectively used for accurate identification of these species. This approach could also be extended to develop specific markers for other Cupressaceae species and their closely related taxa, enabling rapid and reliable identification.

5. Conclusion

Through comprehensive chloroplast genome analyses of ten Cupressaceae species, this study substantially enriches the chloroplast genomic resources of the family. It elucidates interspecific variations in sequence composition, structural organization, and codon usage bias. Notably, highly variable regions within the *rpoA* and *rpoC2* genes were identified, from which multiple diagnostic SNP markers were developed, enabling accurate species discrimination. Given the important ecological role of *C. funebris* and its considerable economic value in the development of cypress essential oils and medicinal products, these findings provide both theoretical and practical foundations for future taxonomic, phylogenetic, and applied research in Cupressaceae.

Authors’ contributions

Sen Cao: Writing – original draft, Methodology, Formal analysis, Data curation. **Zhen Zhang:** Writing – review & editing, Investigation, Funding acquisition, Conceptualization. **Zhichun Zhou:** Supervision, Project administration, Conceptualization. **Haizhen Zhang:** Investigation, Data curation. **Xiuli Chu:** Investigation, Data curation. **Ting Wang:** Methodology,

Investigation, Data curation. **Junzhe Gu:** Methodology, Investigation, Data curation.

Funding

This work was supported by the Major Scientific and Technological Project of Zhejiang Province (No. 2021C02070-8)

Data availability

Sequence data that support the findings of this study have been deposited in the NCBI Genbank database (https://www.ncbi.nlm.nih.gov) with the primary accession code PX355009-PX355018.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Sugiura, M., 1989. The chloroplast chromosomes in land plants. Annu. Rev. Cell Biol. 5, 51–70.
2. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322, 572-574. https://doi.org/10.1038/322572a0.
3. De Las Rivas, J., Lozano, J.J., Ortiz, A.R., 2002. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. Genome Res. 12, 567-583. https://doi.org/10.1101/gr.209402.
4. Kato, T., Kaneko, T., Sato, S., Nakamura, Y., Tabata, S., 2000. Complete Structure of the Chloroplast Genome of a Legume, *Lotus japonicus*. DNA Res. 7, 323-330. https://doi.org/10.1093/dnares/7.6.323.
5. Ebert, D., Peakall, R., 2009. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol. Eco. Resour. 9, 673-690. https://doi.org/10.1111/j.1755-0998.2008.02319.x.
6. Chen, H.M., Shao, J.J., Zhang, H., Jiang, M., Huang, L.F., Zhang, Z., Yang, D., He, M., Ronaghi, M., Luo, X., 2018. Sequencing and analysis of *Strobilanthes*

cusia (Nees) Kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. Front. Plant Sci. 9, 324. https://doi.org/10.3389/fpls.2018.00324.

7. Jiang, M., Chen, H., He, S., Wang, L., Chen, A.J., Liu, C., 2018. Sequencing, Characterization, and Comparative Analyses of the Plastome of *Caragana rosea* var. *rosea*. Int. J. Mol. Sci. 19, 1419. https://doi.org/10.3390/ijms19051419.
8. Liu, X., Zhou, B., Yang, H., Li, Y., Yang, Q., Lu, Y., Gao, Y., 2018. Sequencing and Analysis of *Chrysanthemum carinatum* Schousb and *Kalimeris indica*. The Complete Chloroplast Genomes Reveal Two Inversions and *rbcL* as Barcoding of the Vegetable. Molecules 23, 1358. https://doi.org/10.3390/molecules23061358.
9. Pang, X., Liu, H., Wu, S., Yuan, Y., Li, H., Dong, J., Liu, Z., An, C., Su, Z., Li, B., 2019. Species Identification of Oaks (*Quercus* L., Fagaceae) from Gene to Genome. Int. J. Mol. Sci. 20, 5940. https://doi.org/10.3390/ijms20235940.
10. Cui, Y., Nie, L., Sun, W., Xu, Z., Wang, Y., Yu, J., Song, J., Yao, H., 2019. Comparative and phylogenetic analyses of ginger (*Zingiber officinale*) in the family Zingiberaceae based on the complete chloroplast genome. Plants 8, 283. https://doi.org/10.3390/plants8080283.
11. Gao, C., Deng, Y., Wang, J., 2019. The Complete Chloroplast Genomes of *Echinacanthus* Species (Acanthaceae): Phylogenetic Relationships, Adaptive Evolution, and Screening of Molecular Markers. Front. Plant Sci. 9, 1989. https://doi.org/10.3389/fpls.2018.01989.
12. Li, B., Li, Y., Cai, Q., Lin, F., Huang, P., Zheng, Y., 2016. Development of chloroplast genomic resources for *Akebia quinata* (Lardizabalaceae). Conserv. Genet. Resour. 8, 447-449. https://doi.org/10.1007/s12686-016-0593-0.
13. Moore, M.J., Soltis, P.S., Bell, C.D., Burleigh, J.G., Soltis, D.E., 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. Proc. Natl. Acad. Sci. 107, 4623-4628. https://doi.org/10.1073/pnas.0907801107.
14. Mabberley, D., 2017. Mabberley's plant-book: a portable dictionary of plants, their classification and uses. Cambridge University Press, Cambridge.
15. Mao, K.S., Milne, R.I., Zhang, L., Peng, Y., Liu, J., Thomas, P., Mill, R.R., S. Renner, S., 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proc. Natl. Acad. Sci. 109, 7793-7798. https://doi.org/10.1073/pnas.1114319109.
16. Yang, Z.Y., Ran, J.H., Wang, X.Q., 2012. Three genome-based phylogeny of Cupressaceae sl: further evidence for the evolution of gymnosperms and Southern Hemisphere biogeography. Mol. Phylogenet. Evol. 64, 452-470. https://doi.org/10.1016/j.ympev.2012.05.004.
17. Farjon, A., 2005. A monograph of Cupressaceae and Sciadopitys. Royal Botanic Gardens Kew, Richmond, Surrey.
18. Jiang, Z.P., Wang, H.R., 1997. Taxonomy of the Cupressaceae: Subfamilies, Tribes and Genera. Journal of university of Chinese academy of Science, 236-

248.
19. Gadek, P.A., Alpers, D.L., Heslewood, M.M., Quinn, C.J., 2000. Relationships within Cupressaceae sensu lato: a combined morphological and molecular approach. Am. J. Bot. 87, 1044-1057. https://doi.org/10.2307/2657004.
20. Li, C.X., Yang, Q., 2002. Divergence time estimates for major lineages of Cupressaceae (s. l.). Acta Phytotaxonomica Sinica 4, 323-333.
21. Adams, R.P., Bartel, J.A., Price, R.A., 2009. A new genus, *Hesperocyparis*, for the cypresses of the Western Hemisphere (Cupressaceae). Phytologia 91, 160-185.
22. Little, D.P., 2006. Evolution and circumscription of the true cypresses (Cupressaceae: *Cupressus*). Syst. Bot. 31, 461-480. https://doi.org/10.1600/036364406778388638
23. Little, D.P., Schwarzbach, A.E., Adams, R.P., Hsieh, C.F., 2004. The circumscription and phylogenetic relationships of Callitropsis and the newly described genus *Xanthocyparis* (Cupressaceae). Am. J. Bot. 91, 1872-1881. https://doi.org/10.3732/ajb.91.11.1872.
24. Yang, H., Zhang, R., Jin, G., Feng, Z., Zhou, Z., 2016. Assessing the Genetic Diversity and Genealogical Reconstruction of Cypress (*Cupressus funebris* Endl.) Breeding Parents Using SSR Markers. Forests 7, 160. https://doi.org/10.3390/f7080160.
25. Wang, R.W., Guo, Y., Lin, H., Liu, Z.F., Zhang, R., Zhou, Z.C., 2021. Research Progress and Prospect of *Cupressus funebris*. Hubei Forestry Science and Technology 50, 60-63.
26. Adams, R.P., Li, S.F., 2008. The botanical source of Chinese cedarwood oil: *Cupressus funebris* or Cupressaceae species? J. Essent. Oil Res. 20, 235-242. https://doi.org/10.1080/10412905.2008.9700001.
27. Mao, K., Ruhsam, M., Ma, Y., Graham, S.W., Liu, J., Thomas, P., Milne, R.I., Hollingsworth, P.M., 2018. A transcriptome-based resolution for a key taxonomic controversy in Cupressaceae. Ann. Bot. 123, 153-167. https://doi.org/10.1093/aob/mcy152.
28. Li, J.L., Wu, D.Y., Mao, K.S., 2024. Taxonomy and diversity of the genus *Cupressus*: current status and recommendations. Journal of Nanjing Forestry University 48, 36-45. https://doi.org/10.12302/j.issn.1000-2006.202405017.
29. Zhu, A., Fan, W., Adams, R.P., Mower, J.P., 2018. Phylogenomic evidence for ancient recombination between plastid genomes of the *Cupressus-Juniperus-Xanthocyparis* complex (Cupressaceae). BMC Evol. Biol. 18, 1-10. https://doi.org/10.1186/s12862-018-1258-2.
30. Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., DePamphilis, C.W., Yi, T.-S., Li, D.-Z., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21, 1-31. https://doi.org/10.1186/s13059-020-02154-5.
31. Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., Liu, C., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47, W65-W73. https://doi.org/10.1093/nar/gkz345.

32. Greiner, S., Lehwark, P., Bock, R., 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47, W59-W64. https://doi.org/10.1093/nar/gkz238.
33. Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. 25, 955-964. https://doi.org/10.1093/nar/25.5.955.
34. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32, W273-W279. https://doi.org/10.1093/nar/gkh458.
35. Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394-1403. https://doi.org/10.1101/gr.2289704.
36. Weber, S., Ramirez, C., Doerfler, W., 2020. Signal hotspot mutations in SARS-CoV-2 genomes evolve as the virus spreads and actively replicates in different parts of the world. Virus Res. 289, 198170. https://doi.org/10.1016/j.virusres.2020.198170.
37. Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R., 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29, 4633-4642. https://doi.org/10.1093/nar/29.22.4633.
38. Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics 33, 2583-2585. https://doi.org/10.1093/bioinformatics/btx198.
39. Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30, 772-780. https://doi.org/10.1093/molbev/mst010.
40. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32, 268-274. https://doi.org/10.1093/molbev/msu300.
41. Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293-W296. https://doi.org/10.1093/nar/29.22.4633.
42. Chen, S.L., Lee, W., Hottes, A.K., Shapiro, L., McAdams, H.H., 2004. Codon usage between genomes is constrained by genome-wide mutational processes. Proc. Natl. Acad. Sci. 101, 3480-3485. https://doi.org/10.1073/pnas.0307827100.
43. Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., Ikemura, T., 2001. Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. Gene 276, 89-99. https://doi.org/10.1016/S0378-1119(01)00673-4.
44. Lynn, D.J., Singer, G.A., Hickey, D.A., 2002. Synonymous codon usage is

subject to selection in thermophilic bacteria. Nucleic Acids Res. 30, 4272-4277. https://doi.org/10.1093/nar/gkf546.
45. Saha, M.C., Cooper, J.D., Mian, M.R., Chekhovskiy, K., May, G.D., 2006. Tall fescue genomic SSR markers: development and transferability across multiple grass species. Theor. Appl. Genet. 113, 1449-1458. https://doi.org/10.1007/s00122-006-0391-2.
46. Arif, I.A., Bakir, M.A., Khan, H.A., Al Farhan, A.H., Al Homaidan, A.A., Bahkali, A.H., Al Sadoon, M., Shobrak, M., 2010. A Brief Review of Molecular Techniques to Assess Plant Diversity. Int. J. Mol. Sci. 11, 2079-2096. https://doi.org/10.3390/ijms11052079.
47. Vieira, M.L.C., Santini, L., Diniz, A.L., Munhoz, C.d.F., 2016. Microsatellite markers: what they mean and why they are so useful. Genet. Mol. Biol. 39, 312-328. https://doi.org/10.1590/1678-4685-GMB-2016-0027
48. Moose, S.P., Mumm, R.H., 2008. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol. 147, 969-977. https://doi.org/10.1104/pp.108.118232.
49. Burke, S.V., Lin, C.S., Wysocki, W.P., Clark, L.G., Duvall, M.R., 2016. Phylogenomics and plastome evolution of tropical forest grasses (*Leptaspis*, *Streptochaeta*: Poaceae). Front. Plant Sci. 7, 01993. https://doi.org/10.3389/fpls.2016.01993.
50. Du, Y.P., Bi, Y., Yang, F.P., Zhang, M.F., Chen, X.Q., Xue, J., Zhang, X.H., 2017. Complete chloroplast genome sequences of *Lilium*: insights into evolutionary dynamics and phylogenetic analyses. Sci. Rep. 7, 5751. https://doi.org/10.1038/s41598-017-06210-2.
51. Terry, R.G., Schwarzbach, A.E., Bartel, J.A., 2018. A molecular phylogeny of the Old World cypresses (*Cupressus*: Cupressaceae): Evidence from nuclear and chloroplast DNA sequences. Plant Syst. Evol. 304, 1181-1197. https://doi.org/10.1007/s00606-018-1540-1.
52. Silba, J., 2005. A monograph of the genus Cupressus L. in the twenty-first century. J Int Conifer Preserv Soc 12, 30-103.
53. Mao, K.S., Hao, G., Liu, J., Adams, R.P., Milne, R.I., 2010. Diversification and biogeography of *Juniperus* (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. New Phytol. 188, 254-272. https://doi.org/10.1111/j.1469-8137.2010.03351.x.
54. TÜRKTAŞ, M., Aslay, M., Kaya, E., ERTUĞRUL, F., 2012. Molecular characterization of phylogeneticrelationships in Fritillaria species inferred from chloroplast trnL-trnF sequences. Turk. J. Biol. 36, 552-560. https://doi.org/10.3906/biy-1201-30.
55. Park, I., Kim, W.J., Yeo, S.-M., Choi, G., Kang, Y.-M., Piao, R., Moon, B.C., 2017. The Complete Chloroplast Genome Sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and Comparative Analysis with Other *Fritillaria Species*. Molecules 22, 982. https://doi.org/10.3390/molecules22060982.
56. Li, Y., Zhang, Z.R., Yang, J.B., Lv, G.H., 2018. Complete chloroplast genome of seven *Fritillaria* species, variable DNA markers identification and

phylogenetic relationships within the genus. PLoS One 13, e0194613. https://doi.org/10.1371/journal.pone.0194613.
57. Chen, Q., Wu, X., Zhang, D., 2019. Phylogenetic analysis of *Fritillaria cirrhosa* D. Don and its closely related species based on complete chloroplast genomes. PeerJ 7, e7480. https://doi.org/10.7717/peerj.7480.
58. Daniell, H., Lin, C.-S., Yu, M., Chang, W.-J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17, 134. https://doi.org/10.1186/s13059-016-1004-2.
59. Redwan, R., Saidin, A., Kumar, S., 2015. Complete chloroplast genome sequence of MD-2 pineapple and its comparative analysis among nine other plants from the subclass Commelinidae. BMC Plant Biol. 15, 196. https://doi.org/10.1186/s12870-015-0587-1.
60. Song, Y., Dong, W., Liu, B., Xu, C., Yao, X., Gao, J., Corlett, R.T., 2015. Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. Front. Plant Sci. 6, 662. https://doi.org/10.3389/fpls.2015.00662.
61. Jheng, C.F., Chen, T.C., Lin, J.Y., Chen, T.C., Wu, W.L., Chang, C.C., 2012. The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish Phalaenopsis orchids. Plant Sci. 190, 62-73. https://doi.org/10.1016/j.plantsci.2012.04.001.
62. Mehmood, F., Ubaid, Z., Shahzadi, I., Ahmed, I., Waheed, M.T., Poczai, P., Mirza, B., 2020. Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). PeerJ 8, e9552. https://doi.org/10.7717/peerj.9552.
63. Chen, C., Xia, X., Peng, J.Y., Wang, D.W., 2022. Comparative analyses of six complete chloroplast genomes from the genus *Cupressus* and *Juniperus* (Cupressaceae). Gene 837, 146696. https://doi.org/10.1016/j.gene.2022.146696.
64. Ping, J., Feng, P., Li, J., Zhang, R., Su, Y., Wang, T., 2021. Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. Ecol. Evol. 11, 4786-4802. https://doi.org/10.1002/ece3.7381.

Figure captions

Fig. 1 Plastmoe map of *C. funebris*. The outermost circle displays the gene map, with genes colorcoded based on their functional categories. The gray inner circle represent GC content across the genome. The name and total size of the genome are shown at the center of the map.

Fig. 2 Comparative analysis of chloroplast genome architectures among 10 Cupressaceae species was conducted using the MAUVE program. The name corresponding to each sequence was shown in the lower left corner. The sequence with the same color represented similar regions across different species, differing only in their relative positions within the genome.

Fig. 3 Compare of the chloroplast genomes among the ten Cupressaceae species with the *C. funebris* as the reference by mVISTA. The genes are illustrated as gray arrows positioned above the alignment, with distinct genomic regions highlighted in different color for clarify. The purple regions are exons, the red regions are “conserved noncoding sequences” (CNS), and the light-blue regions are tRNAs or rRNAs. The percentages 50 and 100% on the right side refer to the similarity among sequences.

Fig. 4 Codon composition corresponding to 21 amino acids across all protein-coding genes in the chloroplast genomes of 10 Cupressaceae species. The histogram of each amino acid indicated codon usage within Cupressaceae (from left to right: *C. funebris*, *C. sempervirens*, *C. pisifera ‘Squarrosa’*, *C. pisifera ‘Filifera’*, *C. obtuse*, *T. dolabrata*, *J. formosana*, *P. orientalis ‘Sieboldii’*, *H. arizonica* and *H. lusitanica*)

Fig. 5 Repeated sequences in 10 Cupressaceae chloroplast genomes.

Fig. 6 Phylogenetic tree inferred from the chloroplast genome sequences of 63 Cupressaceae species. Bootstrap support values are indicated at each node. The 10 Cupressaceae species analyzed in this study were labeled in red bont.

Fig. 7 Partial sequence alignment of genes in the chloroplast genomes of 10 Cupressaceae species. a, gene *rpoA* ; b, gene *rpoC2*. The red arrows indicate SNPs.

Fig. 8 *C. sempervirens* sampled in Shanghai Botanical Garden. a, the whole tree. b, leaves.

Table 1 Comparative analyses on the basic features of the plastomes of 10 Cupressaceae species

	Cupressus funebris	*Cupressus sempervirens*	*Chamaecyparis pisifera 'Squarrosa'*	*Chamaecyparis pisifera 'Filifera'*	*Chamaecyparis obtusa*	*Thujopsis dolabrata*	*Juniperus formosana*	*Platycladus orientalis 'Sieboldii'*	*Hesperocyparis arizonica*	*Hesperocyparis lusitanica*
Length(bp)	127816	128082	127490	127653	127730	129175	128155	127074	127174	127107
GC content(%)	34.71	34.96	35.01	34.98	34.98	34.68	34.89	34.74	34.67	34.68
Gene number	117	117	117	117	117	117	115	118	116	116
Protein-coding gene number	82	82	83	83	83	83	82	83	82	82
rRNA gene number	4	4	4	4	4	4	4	4	4	4
tRNA gene number	31	31	30	30	30	30	29	31	30	30

Table 3 The number of six major SSR types across 10 Cupressaceae speices

	Amount										
SSR type	*C. funebris*	*C. sempervirens*	*C. pisifera 'Squarrosa'*	*C. pisifera 'Filifera'*	*C. obtusa*	*T. dolabrata*	*J. formosana*	*P. orientalis 'Sieboldii'*	*H. arizonica*	*H. lusitanica*	Total
Monomer	40	33	47	46	32	23	36	40	47	42	386
Dimer	10	12	18	22	19	24	7	11	14	13	150
Trimer	6	4	5	5	11	7	8	15	6	8	75
Tetramer	11	5	8	9	8	9	8	9	8	8	83
Pentamer	-	-	1	1	-	-	1	3	-	1	7
Hexamer	1	-	3	3	-	-	1	-	-	-	8

Total	68	54	82	86	70	63	61	78	75	72	709

Table 2 Gene List of Chloroplast Genome in 10 Cupressaceae Species

Group of genes	Name of genes
Subunits of NaDH-dehydrogenase	*ndh*A, *ndh*B, *ndh*C, *ndh*D, *ndh*E, *ndh*F, *ndh*G, *ndh*H, *ndh*I, *ndh*J, *ndh*K
Subunits of photosystem I	*psa*A, *psa*B, *psa*C, *psa*I, *psa*J, *psa*M
Subunits of photosystem Ⅱ	*psb*A, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*I, *psb*J, *psb*K, *psb*L, *psb*M, *psb*N, *psb*T, *psb*Z, *ycf*3
Subunits of cytochrome b/f complex	*pet*A, *pet*B, *pet*D, *pet*G, *pet*L, *pet*N,
Subunits of ATP synthase	*atp*A, *atp*B, *atp*E, *atp*F, *atp*H, *atp*I
Large subunit of rubisco	*rbc*L
Small subunit of ribosome	*rps*2, *rps*3, *rps*4, *rps*7, *rps*8, *rps*11, *rps*12, *rps*14, *rps*15, *rps*16[3,4,5,6,8], *rps*18, *rps*19
Large subunit of ribosome	*rpl*2, *rpl*14, *rpl*16, *rpl*20, *rpl*22, *rpl*23, *rpl*32, *rpl*33, *rpl*36
DNA dependent RNA polymerase	*rpo*A, *rpo*B, *rpo*C1, *rpo*C2
rRNA genes	*rrn*4.5S, *rrn*5S, *rrn*16S, *rrn*23S
tRNA genes	*trn*A-UGC, *trn*C-GCA, *trn*D-GUC, *trn*E-UUC, *trn*F-GAA, *trnf*M-CAU[+], *trn*G-GCC, *trn*G-UCC, *trn*H-GUG, *trn*I-CAU, *trn*I-GAU, *trn*K-UUU, *trn*L-CAA[&], *trn*L-UAA, *trn*L-UAG, *trn*M-CAU, *trn*N-GUU, *trn*P-GGG[*], *trn*P-UGG, *trn*Q-UUG, *trn*R-ACG, *trn*R-UCU, *trn*S-GCU, *trn*S-GGA, *trn*S-UGA, *trn*T-GGU[1,2,6,7,8,9,10], *trn*T-UGU, *trn*V-GAC, *trn*V-UAC, *trn*W-CCA, *trn*Y-GUA
Maturase	*mat*K
c-type cytochrome synthesis gene	*ccs*A
Envelope membrane protein	*cem*A
Protease	*clp*P
Subunit of Acetyl-CoA-carboxylase	*acc*D
Translational initiation factor	*inf*A
Protochlorophyllide reductase	*chl*B, *chl*L, *chl*N
Gene of unknown functions Open Reading	*ycf*1, *ycf*2, *ycf*4

Notes: 1, the gene presented in *C. funebris*. **2**, the gene presented in *C. sempervirens*. **3**, the gene presented in *C. pisifera 'Squarrosa'*. **4**, the gene presented in *C. pisifera 'Filifera'*. **5**, the gene presented in *C. obtuse*. **6**, the gene presented in *T. dolabrata*. **7**, the gene presented in *J. formosana*. **8**, the gene presented in *P. orientalis 'Sieboldii'*. **9**, the gene presented in *H. arizonica.* **10**, the gene presented in *H. lusitanic.* **&**, the gene was absent

only in *J. formosana*. *, the gene was absent only in *T. dolabrata* and *J. formosana.* **+**, the gene was absent only in *H. arizonica* and *H. lusitanic*.

Figures

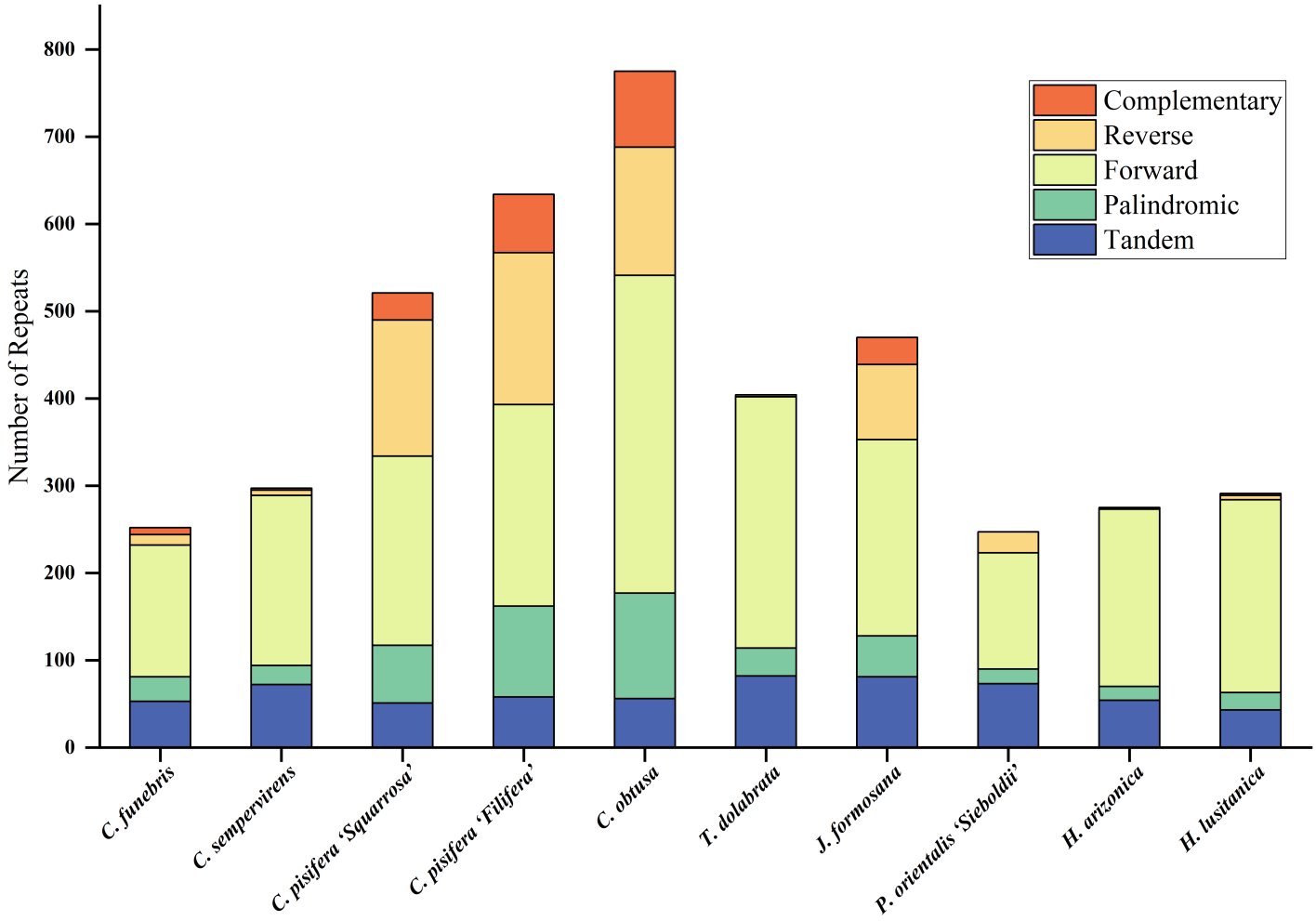


Figure 4

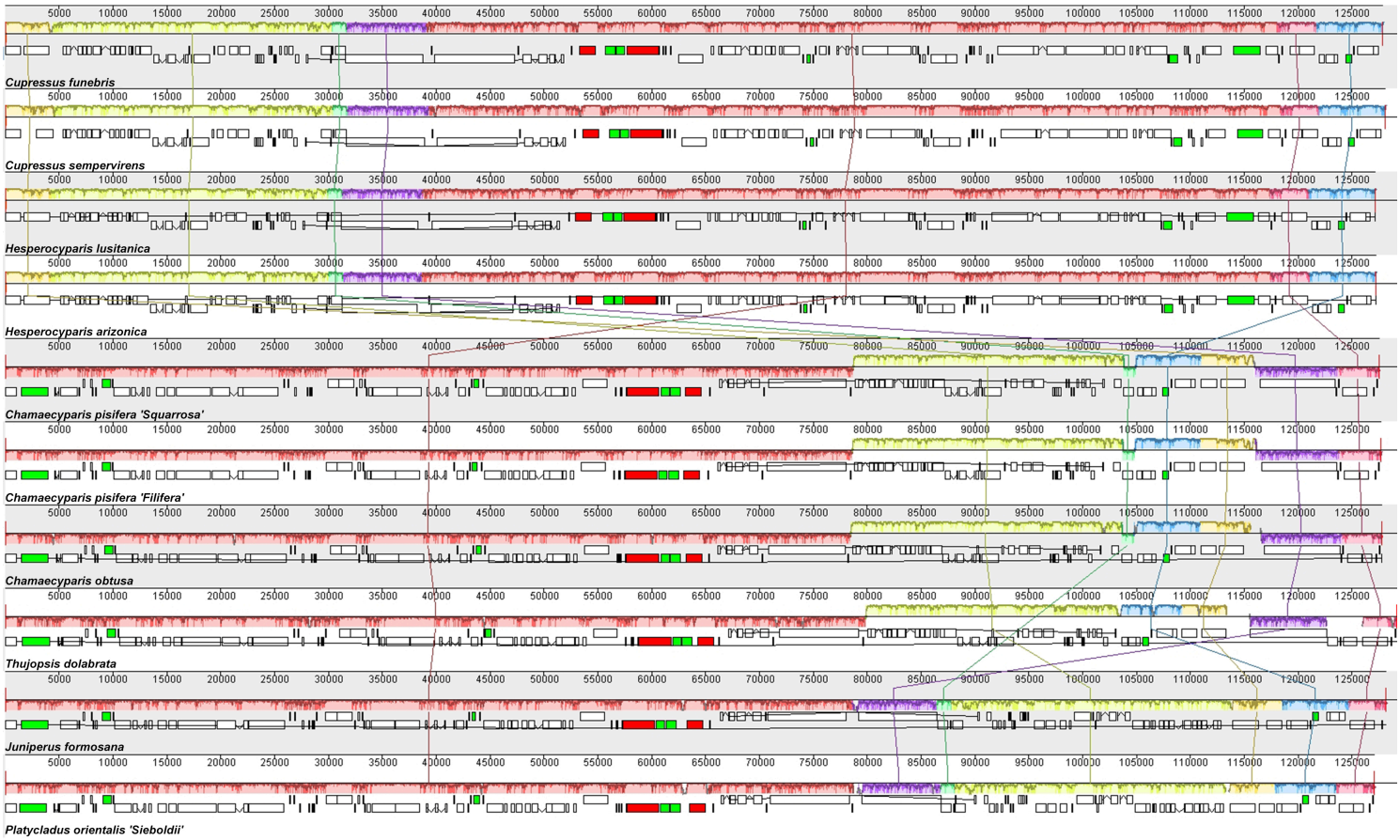


Figure 5



Figure 6



Figure 7



Figure 8



Figure 9



Figure 10



Figure 11

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable3.xlsx
- SupplementaryTable5.xlsx
- SupplementaryTable4.xlsx
- SupplementaryTable12.docx