



## OPEN ACCESS

## EDITED BY

Jia-Yu Xue,  
Nanjing Agricultural University, China

## REVIEWED BY

Zhanyuan Lu,  
Inner Mongolia University, China  
Guang-Can Zhou,  
Heze University, China

## \*CORRESPONDENCE

Fan Zhang,  
✉ mumizhongfeng@126.com  
Xiaoqing Lu,  
✉ luxiaoqing@jib.ac.cn

RECEIVED 24 March 2023

ACCEPTED 27 June 2023

PUBLISHED 02 August 2023

## CITATION

Yue M, Chen H, Xuan L, Yang Y, Chong X,  
Li M, Yu C, Lu X and Zhang F (2023), Novel  
molecular markers for *Taxodium*  
breeding from the chloroplast genomes  
of four artificial *Taxodium* hybrids.  
*Front. Genet.* 14:1193023.  
doi: 10.3389/fgene.2023.1193023

## COPYRIGHT

© 2023 Yue, Chen, Xuan, Yang, Chong, Li,  
Yu, Lu and Zhang. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original author(s)  
and the copyright owner(s) are credited  
and that the original publication in this  
journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Novel molecular markers for *Taxodium* breeding from the chloroplast genomes of four artificial *Taxodium* hybrids

Minghua Yue<sup>1,2</sup>, Hong Chen<sup>1</sup>, Lei Xuan<sup>3</sup>, Ying Yang<sup>3</sup>,  
Xinran Chong<sup>1</sup>, Mingzhi Li<sup>4</sup>, ChaoGuang Yu<sup>3</sup>, Xiaoqing Lu<sup>1\*</sup> and  
Fan Zhang<sup>1\*</sup>

<sup>1</sup>Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China, <sup>2</sup>Nanjing Botanical Garden Mem. Sun Yat-Sen, Nanjing, China, <sup>3</sup>Jiangsu Key Laboratory for the Research and Utilization of Plant Resources, Nanjing, China, <sup>4</sup>Guangzhou Bio&Data Technology Co., Ltd., Guangzhou, China

*Taxodium* “Zhongshanshan” are a group of intraspecific *Taxodium* hybrids with superparental dominance and high ecological and economic value in southern China. Identifying the parentage of hybrids, especially the male parent, is critically important for genetic studies. However, the large nuclear genomes of members of the genus *Taxodium* pose a major challenge for the development of molecular markers. Here, we developed novel molecular markers by conducting a comparative analysis of the chloroplast genomes of four artificial *Taxodium* hybrids and their parents. The lengths of the whole chloroplast genome ranged from 131,942 to 132,128 bp, and the total guanine (GC) content of the chloroplast genomes ranged from 34.6% to 35.81%. A total of 120 unique genes were identified, including 83 protein-coding genes, 33 transfer RNAs, and four ribosomal RNAs. There were 69–71 simple sequence repeats were detected in the four hybrids. Phylogenetic analysis revealed that these hybrids clustered with their paternal parents. Similar findings were obtained by analysis of the GC content of protein-coding genes. Molecular markers were developed using the highly variable regions of the chloroplast genomes, and polymerase chain reaction (PCR) assays revealed that these markers were effective for identifying the male parents of these hybrids. Our findings indicate for the first time that the chloroplast genomes of *Taxodium* are paternally inherited. Generally, these molecular markers could facilitate breeding and genetic studies of *Taxodium*.

## KEYWORDS

*Taxodium*, chloroplast genome, phylogenetic relationships, molecular markers, breeding

## 1 Introduction

The three members of the genus *Taxodium*, *Taxodium distichum* (L.) Rich (bald cypress), *T. mucronatum* Tenore (Montezuma cypress), and *T. ascendens* Brongn (pond cypress), are 30–40 m tall trees native to North America (Denny and Arnold, 2007). *T. distichum* is the most tolerant of flooding and strong winds; *T. mucronatum* is the most tolerant of saline and alkaline soils but is the most susceptible to red blight disease (Zhang et al., 2021); and *T. ascendens* is the most resistant to red blight disease but the least tolerant of flooding. All three of these species have been introduced to southeastern China for their ornamental and ecological value (Wang et al., 2017).

TABLE 1 List of genes in the cp genomes of four *T. 'Zhongshanshan'* in this study.

Category	Gene group	Name of gene	Number
Self-replication	Proteins of the large ribosomal subunit	<i>rpl2<sup>a</sup>, rpl14, rpl16<sup>a</sup>, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>	9
	Proteins of the small ribosomal subunit	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12<sup>b</sup>, rps14, rps15, rps16<sup>a</sup>, rps18, rps19</i>	12
	Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1<sup>a</sup>, rpoC2</i>	4
	rRNAs	<i>rrn23S, rrn16S, rrn5S, rrn4.5S</i>	4
	tRNAs	<i>trnH-GUG, trnK-UUU<sup>a</sup>, trnQ-UUG<sup>c</sup>, trnS-GCU, trnG-UCC, trnR-UCU, trnC-GCA, trnD-GUC, trnY-GUA, trnE-UUC, trnT-GGU, trnS-UGA, trnM-CAU, trnS-GGA, trnT-UGU, trnL-UAA<sup>a</sup>, trnF-GAA, trnV-UAC<sup>c</sup>, trnM-CAU, trnW-CCA, trnP-UGG, trnI-CAU<sup>c</sup>, trnV-GAC, trnI-GAU<sup>a</sup>, trnL-CAA, trnA-UGC<sup>c</sup>, trnR-ACG, trnN-GUU, trnL-UAG, trnP-GGG, trnG-GCC</i>	33
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psam, psal, psaj, ycf3<sup>b</sup>, ycf4</i>	8
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
	Subunits of NADH NADH-plastoquinone oxidoreductase	<i>ndhA<sup>a</sup>, ndhB<sup>a</sup>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	11
	Subunits of cytochrome b/f complex	<i>petA, petB<sup>a</sup>, petD<sup>a</sup>, petG, petL, petN</i>	6
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF<sup>a</sup>, atpH, atpI</i>	6
	Large subunit of Rubisco	<i>rbcl</i>	1
	Subunits of photochlorophyllide reductase	<i>chlB, chlL, chlN</i>	3
Biosynthesis	Maturase	<i>matK</i>	1
	Protease	<i>clpP</i>	1
	Envelope membrane protein	<i>cemA</i>	1
	Acetyl-CoA carboxylase	<i>accD</i>	1
	c-type cytochrome synthesis gene	<i>ccsA</i>	1
	Translation initiation factor	<i>infA</i>	1
Unknown function	Conserved hypothetical chloroplast reading frames	<i>ycf1, ycf2</i>	2

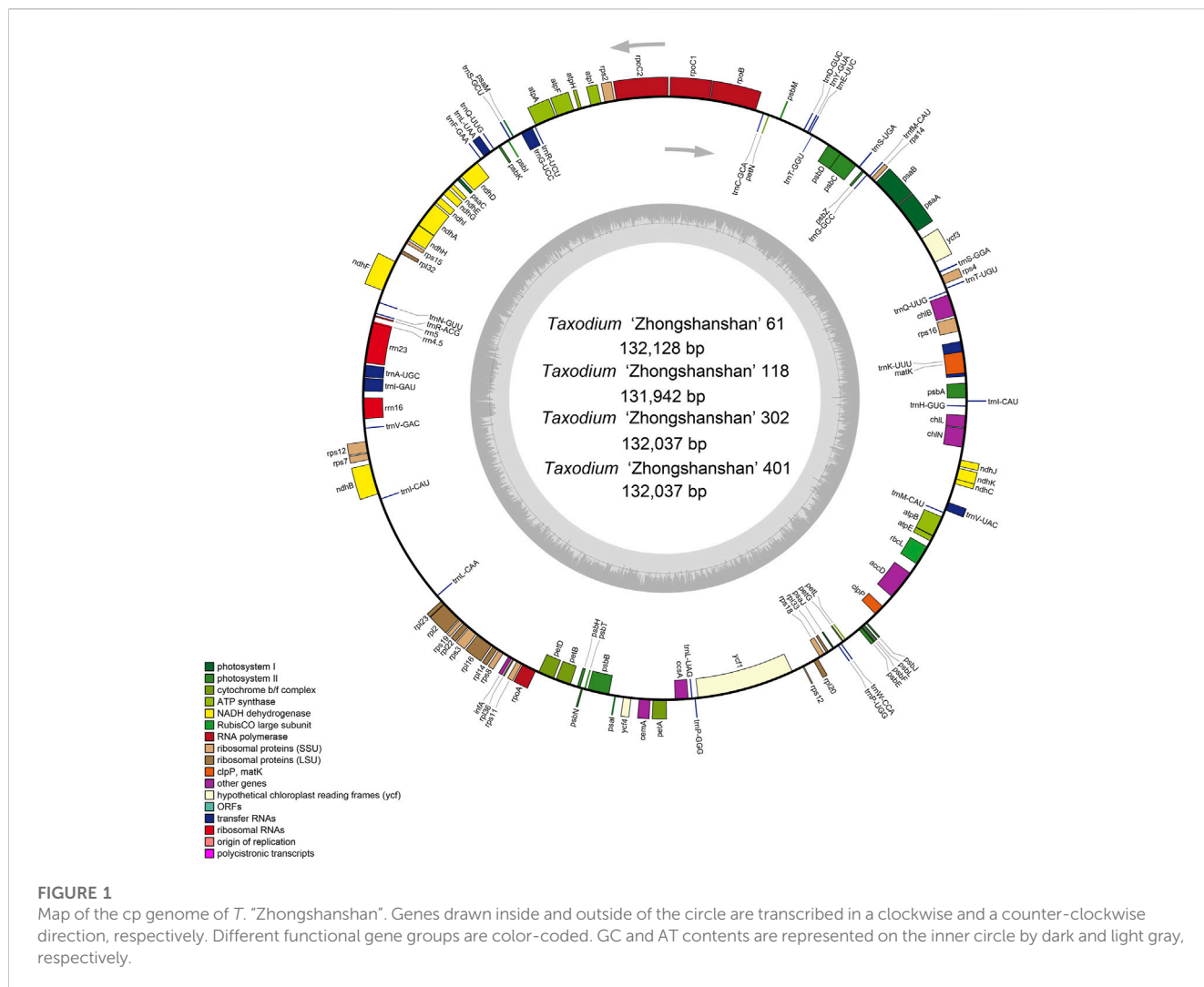
<sup>a</sup>Genes containing two exons; <sup>b</sup> Genes containing three exons; <sup>c</sup> Two copies.

Interspecific hybridization can result in the production of new species; additionally, it can be used to generate hybrids with desirable characteristics of species crossed. Previously, we artificially hybridized the three *Taxodium* species to generate several new varieties, which were referred to as “Zhongshanshan”; these were selected from the progeny because of their superior features (Crech et al., 2011). For example, the desirable characteristics of *T. “Zhongshanshan 302”* (*T. distichum* × *T. mucronatum*) and *T. “Zhongshanshan 401”* (*T. ascendens* × *T. mucronatum*) included their high growth rate, ability to grow well in alkaline and saline soils, and high ornamental value (Wang et al., 2016). The growth rate of *T. “Zhongshanshan 118”* (*T. “Zhongshanshan 302”* × *T. mucronatum*) is higher than that of its parents; the salt, alkali, and flood resistance of *T. “Zhongshanshan 118”* is also higher than that of its parents. Because of the desirable characteristics of these varieties, they have become increasingly popular in 18 provinces and municipalities in China (Xuan et al., 2021). Previously, expressed sequence tag-simple sequence repeat markers identified via transcriptome sequencing, genetic linkage maps built using sequence-related amplified polymorphism markers (Cheng et al., 2015), simple sequence repeat (SSR) markers (Wang et al., 2016), and high-density genetic maps built using specific locus amplified fragment sequencing technology (Yang et al., 2018) have greatly aided *Taxodium* breeding.

Comprehensive genomic datasets of the genus *Taxodium* are needed to facilitate genetic studies and breeding. However, the large size of the nuclear genomes of members of the genus *Taxodium* makes generating these datasets a major challenge.

The chloroplast (cp) is a critically important organelle in green plants; it contains the photosynthetic machinery that plays a key role in carbon fixation, as well as plant growth and development (Palmer, 1987). The rate of evolutionary change of cp genomes is slower than that of nuclear genomes; genes and genome organization are more highly conserved in cp genomes than in nuclear genomes (Androsiuk et al., 2020). In addition, cp genomes are uniparentally inherited. Cp genomes have become increasingly used in phylogenetic analyses, as well as studies aimed at clarifying complex evolutionary relationships. Cp genome sequences have also been used for the development of molecular markers that can be used to distinguish morphologically similar species.

Here, we characterized the complete cp genomes of four *T. “Zhongshanshan”* varieties and compared them with the cp genome sequences of their parents. Phylogenetic analysis and analysis of the GC content of protein-coding genes revealed that these hybrids were most closely related to their male parents. Analysis of hypervariable regions was used to generate molecular markers that could be used to aid *Taxodium* breeding.



## 2 Materials and methods

### 2.1 Plant material and chloroplast genome sequencing, assembly, and annotation

The superior clones of *T. "Zhongshanshan 302"* (*T. distichum* × *T. mucronatum*), *T. "Zhongshanshan 118"* (*T. "Zhongshanshan 302"* × *T. mucronatum*), *T. "Zhongshanshan 61"* (*T. mucronatum* × *T. ascendens*), and *T. "Zhongshanshan 401"* (*T. ascendens* × *T. mucronatum*) were grown in a nursery at the Nanjing Botanical Garden (35°50' N, 45°70' E), Jiangsu Province, China. Fresh leaves of these plant were collected for DNA extraction with a modified CTAB method (Doyle and Doyle, 1987). DNA quality was assessed in a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, United States), and its integrity and concentration were evaluated using 1% gel electrophoresis and a Qubit fluorometer (Life Technologies, Darmstadt, Germany), respectively.

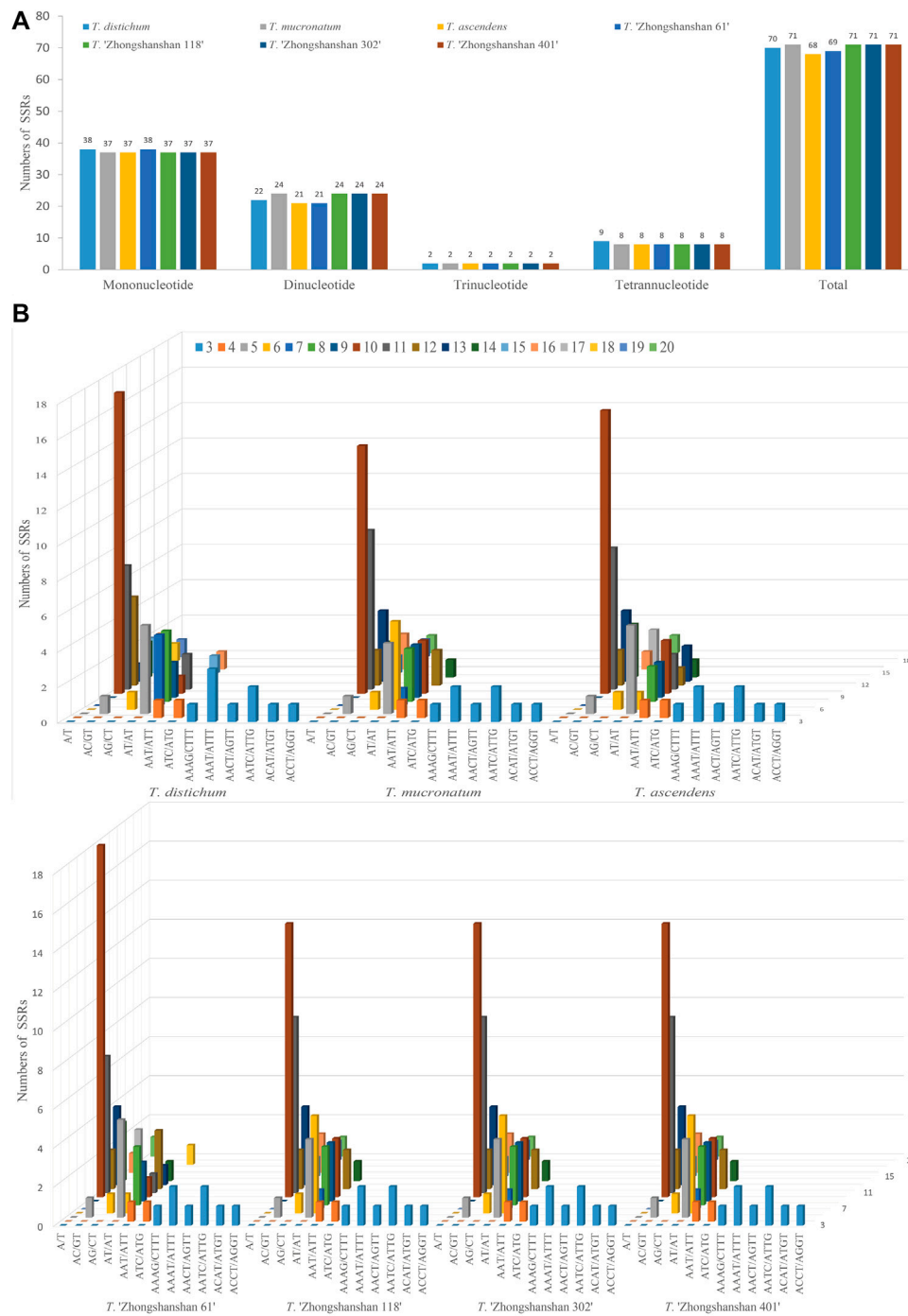
The DNA library was sequenced by BGISEQ-500 platform (Bio&Data Biotechnologies Co. Ltd., Nanjing, China) with 150 based on sequencing by synthesis. Approximately 43 Mb of high-quality, clean paired-end reads was generated, and the

sequences with cp-like reads were assembled with NOVOPlasty Version 4.3.1 (Dierckxsens et al., 2017).

The genome was annotated using GeSeq software (Tillich et al., 2017) and BLAST, and tRNAs were identified by tRNA scan-SE (Schattner et al., 2005). The results were further checked manually, and the circular graphical maps of the entire genome were drawn using the OGDRAW v1.3.1 program (<http://ogdraw.mpimp-golm.mpg.de>) (Lohse et al., 2007). The cp genome sequences of *T. "Zhongshanshan"* were deposited in the GenBank of NCBI (<https://www.ncbi.nlm.nih.gov/>) database under the accession numbers MW307789, MW307790, MW307791, and MW307792. The associated bioproject numbers are SRR13347061, SRR13347063, and SRR13347064.

### 2.2 Polymerase chain reaction assays

The highly variable regions were amplified with the high-fidelity polymerase of KOD-PlusNeo (TaKaRa, Dalian, China). The PCR reactions were performed in a total volume of 50 μL with 5 μL 10 × KOD-plus Neo buffer, 5 μL dNTP (2 mM), 3 μL MgSO<sub>4</sub> (25 mM), 1.5 μL per primers (10 μM), 1 μL KOD-Plus Neo enzyme, and 100 ng genomic DNA. Thermal cycling

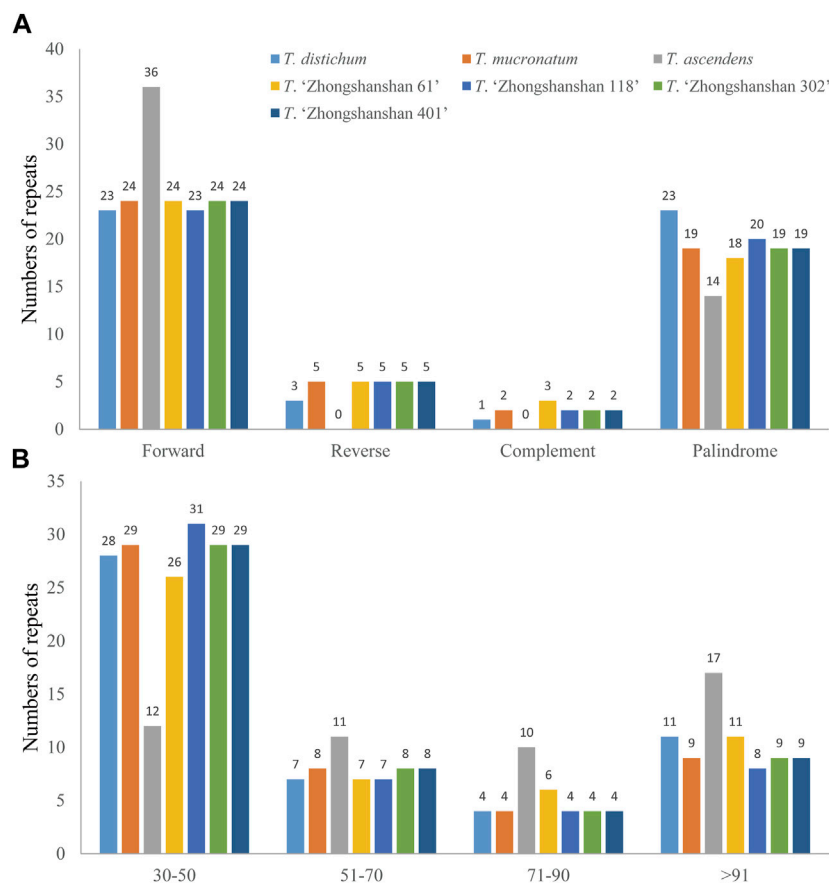


**FIGURE 2** Analysis of SSRs in the cp genomes of eight *Taxodium* species. Number of different SSRs types (A) and SSR motifs (B) in the cp genomes of *Taxodium*.

consisted of 94°C for 2 min followed by 35 cycles of 98°C for 20 s, 62°C for 30 s, and 72 °C for 30 s and a final extension at 72 °C for 7 min 2.5% agarose gels were used to visualize the PCR products after electrophoresis.

### 2.3 Repeat sequence analysis and codon usage analysis

Simple sequence repeats (SSRs) were identified by MISA v2.1 (<https://webblast.ipk-gatersleben.de/misa>) software (Mudunuri and Nagarajaram, 2007) with the following parameters of mono-, di-, tri-, tetra-, penta-, and hexanucleotide being set as a minimum number of repeats of 10, 5, 4, 3, 3, and 3, respectively. The long repeat sequences,



**FIGURE 3** Comparison of long repeats in the cp genomes of seven *Taxodium* species. (A) Number of different types of long repeats. (B) Number of repeats in each repeat-length type.

which included forward, reverse, palindrome, and complement repeats were analyzed by the REPuter (<https://bibiserv.Cebitec.uni-bielefeld.de/reputer>) software with the best top 50 sequences (Kurtz et al., 2001). The CodonW1.4.2 program was used to analyze the RSCU base on the protein-coding genes.

## 2.4 Chloroplast genome comparison

The mVISTA (<http://genome.lbl.gov/vista/mvista>) (Mayor et al., 2000) program was used to examine the genetic divergence among eight complete cp genomes, consisting of *T. distichum* (NC\_034941), *T. distichum* (MN535013), *T. "Zhongshanshan 61"* (MW307790), *T. ascendens* (MN535012), *T. "Zhongshanshan 118"* (MW307791), *T. "Zhongshanshan 302"* (MW307792), *T. "Zhongshanshan 401"* (MW307789), and *T. mucronatum* (MN535011), in the ShuffleLAGAN mode. BLAST Atlas on the GView server was used to assess the chloroplast genomes similarity with *T. distichum* (MN535013) genome as a reference (Petkau et al., 2010). DNAsP v5.0 software was used to calculate nucleotide diversity (Pi) among the eight *Taxodium* cp genomes. Pi values were calculated in 100 bp sliding windows with 25 bp steps. EMBOSS software was used to calculate the GC content at the first, second, and third codon positions (GC1, GC2, and GC3, respectively) and overall GC content of the genomes (Rice et al., 2000).

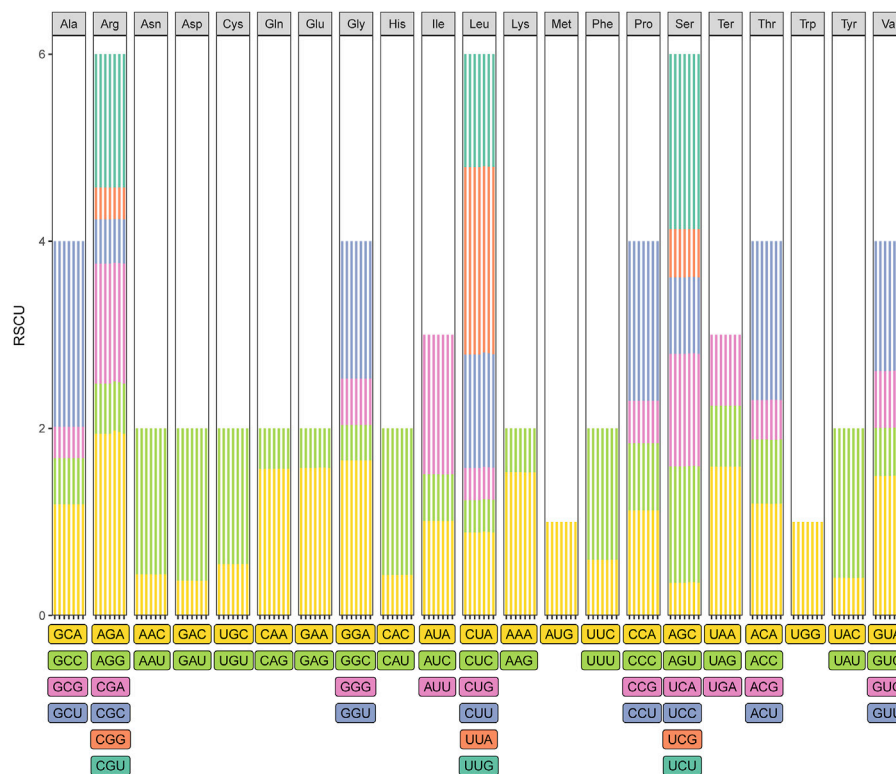
## 2.5 Phylogenetic analysis

For phylogenetic analysis, the whole cp genomes of eight *Taxodium* and an outgroup (*Cryptomeria japonica*) were downloaded from the National Center for Biotechnology Information (NCBI) database. MAFFT software (v7.450) was used to multiply sequence alignment of these cp genome sequences (Kato et al., 2019). The maximum likelihood (ML) method based complete cp genome in IQ-TREE v2.1.4 was conducted for phylogenetic analysis with default parameter settings and 1,000 bootstrap replicates (Minh et al., 2020).

## 3 Results and discussion

### 3.1 Features of the *T. "Zhongshanshan"* cp genome

A total of 5.17–6.52 Gb of clean data of complete cp genomes were obtained from the four *T. "Zhongshanshan"* hybrids using the BGISEQ-500 platform (Supplementary Table S1). The cp genomes were *de novo* assembled using NOVOPlasty, and the length of the whole cp genome ranged from 131,942 bp for *T. "Zhongshanshan 118"* to 132,128 bp for *T. "Zhongshanshan 61"* (Supplementary Table S2).



**FIGURE 4**

Codon usage in the *Taxodium* cp genomes. Different colors correspond to codons listed underneath the columns. Each histogram from left to right was *T. mucronatum*, *T.* “Zhongshanshan 401”, *T.* “Zhongshanshan 302”, *T.* “Zhongshanshan 118”, *T. ascendens*, *T.* “Zhongshanshan 61”, and *T. distichum*, respectively.

The cp genome contains four characteristic regions: a large single-copy region (LSC) and a small single-copy region (SSC), which is separated by two inverted repeat (IRs) regions. Previous studies have shown that IRs play key roles determining the stability of the genome and the conservation of genes (Ping et al., 2022). However, the IR region is lacking in some plant families such as Pinaceae and Taxaceae (Yi et al., 2013). IR regions are absent in the cp genomes of the four hybrids (Figure 1), which is consistent with analyses of the cp genomes of their parents (Duan et al., 2020).

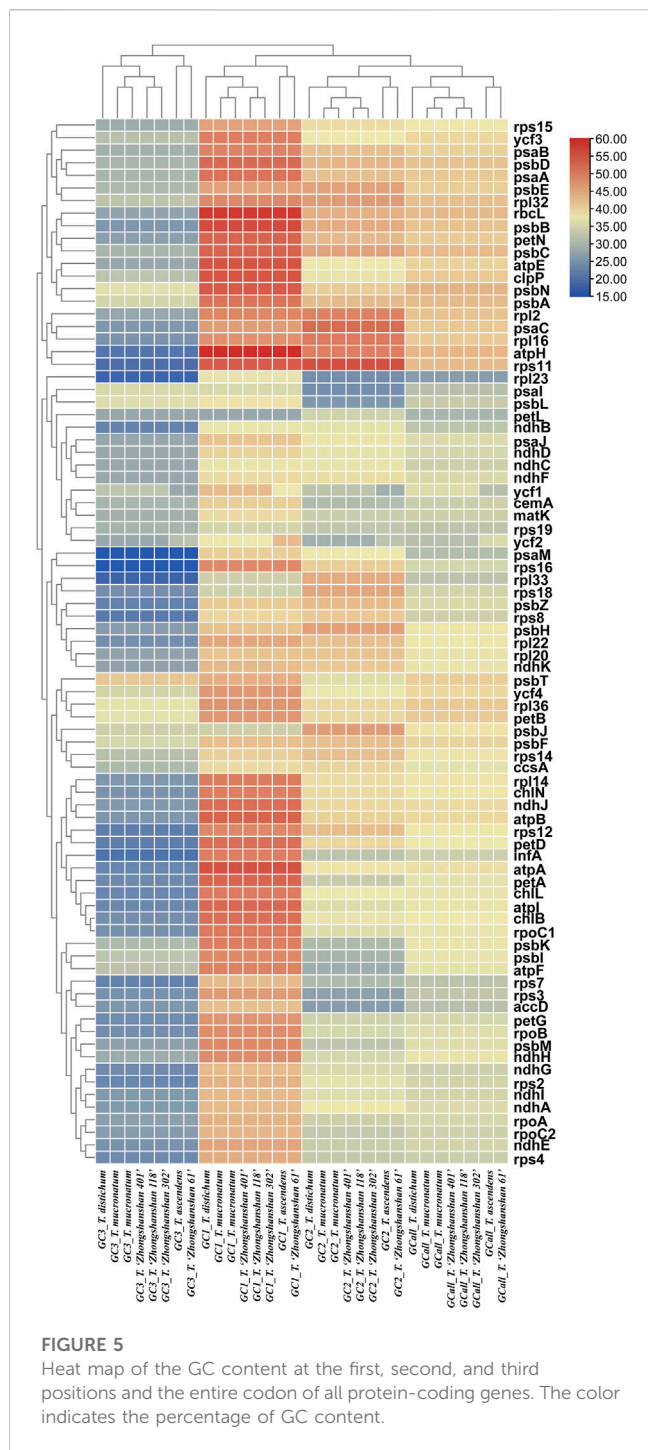
All four *T.* “Zhongshanshan” cp genomes contained 120 genes, including 83 protein-coding genes, 33 tRNAs, and four rRNAs; the distribution of these genes was consistent among all four “Zhongshanshan” cp genomes examined. A total of 62 genes belonged to the self-replication category; 33 of these encoded tRNAs, 12 of these encoded small ribosomal subunit proteins, and nine of these encoded large ribosomal subunit proteins. Fifty of these genes were related to photosynthesis, including six genes that encoded ATP synthase subunits, 11 genes that encoded NADH-plastoquinone oxidoreductase subunits, six genes that encoded cytochrome b/f complex subunits, eight genes that encoded photosystem I subunits, 15 genes that encoded PSII subunits, three genes that encoded photochlorophyllide reductase subunits, and one gene that encoded the large subunit of Rubisco. A total of six groups were in the biosynthesis category, and each group contained one gene (Table 1). A total of 17 intron-containing genes were identified in the *T.* “Zhongshanshan” cp genome; 15 of these genes contained two exons, and two of these

genes (*ycf3* and *rps12*) contained three exons. Two copies of the *trnQ-UUG* and *trnI-CAU* genes were detected in the cp genomes of *T.* “Zhongshanshan” (Table 1). The AT content of the cp genome was 64.7%, which indicates that the nucleotide composition was AT-biased.

### 3.2 Analysis of SSRs and long repeats

To clarify differences between hybrid and parental cp genomes, a comparative analysis of the cp genomes of four new *T.* “Zhongshanshan” hybrids and their parents (*Taxodium distichum*, *T. mucronatum*, and *T. ascendens*) was conducted. 68–71 SSRs were detected in the cp genomes (Figure 2A). The most common motifs were mononucleotide SSRs (37–38, 52.11%–55.07%); mononucleotide A was the most common, followed by mononucleotide T (Supplementary Table S3). The number of trinucleotide and tetranucleotide repeats was consistent among the seven *Taxodium* cp genomes, with the exception of the cp genome of *T. distichum*, which had nine tetranucleotide repeats. The AT/AT sequence of dinucleotide SSRs was the most common (19–22, 27.53%–31.99%), followed by AC/GT (1, 0.01%) and AG/CT (1, 0.01%) (Figures 2B,C).

Four types of long repeats (forward, palindromic, reverse, and complementary) were identified in the cp genomes of the seven *Taxodium* species examined. Forward repeats were the most common type in the cp genomes, followed by palindromic repeats (Figure 3A). Reverse and complementary repeats were



absent in the cp genomes of *T. ascendens*, and five reverse repeats (with the exception of *T. distichum*, which has three reverse repeats) and two complementary repeats (except for *T. distichum* and *T.* “Zhongshanshan 61”, which have one and three complement repeats, respectively), were identified in the other species (Figure 3A; Supplementary Table S4). In addition, sequence lengths of 30–50 bp were the most common in these cp genomes, especially in *T.* “Zhongshanshan 118” (31, 62%), followed by sequence lengths of >91 bp, then 51–70 bp and 71–90 bp (Figure 3B). The distribution of sequence lengths was even in *T.*

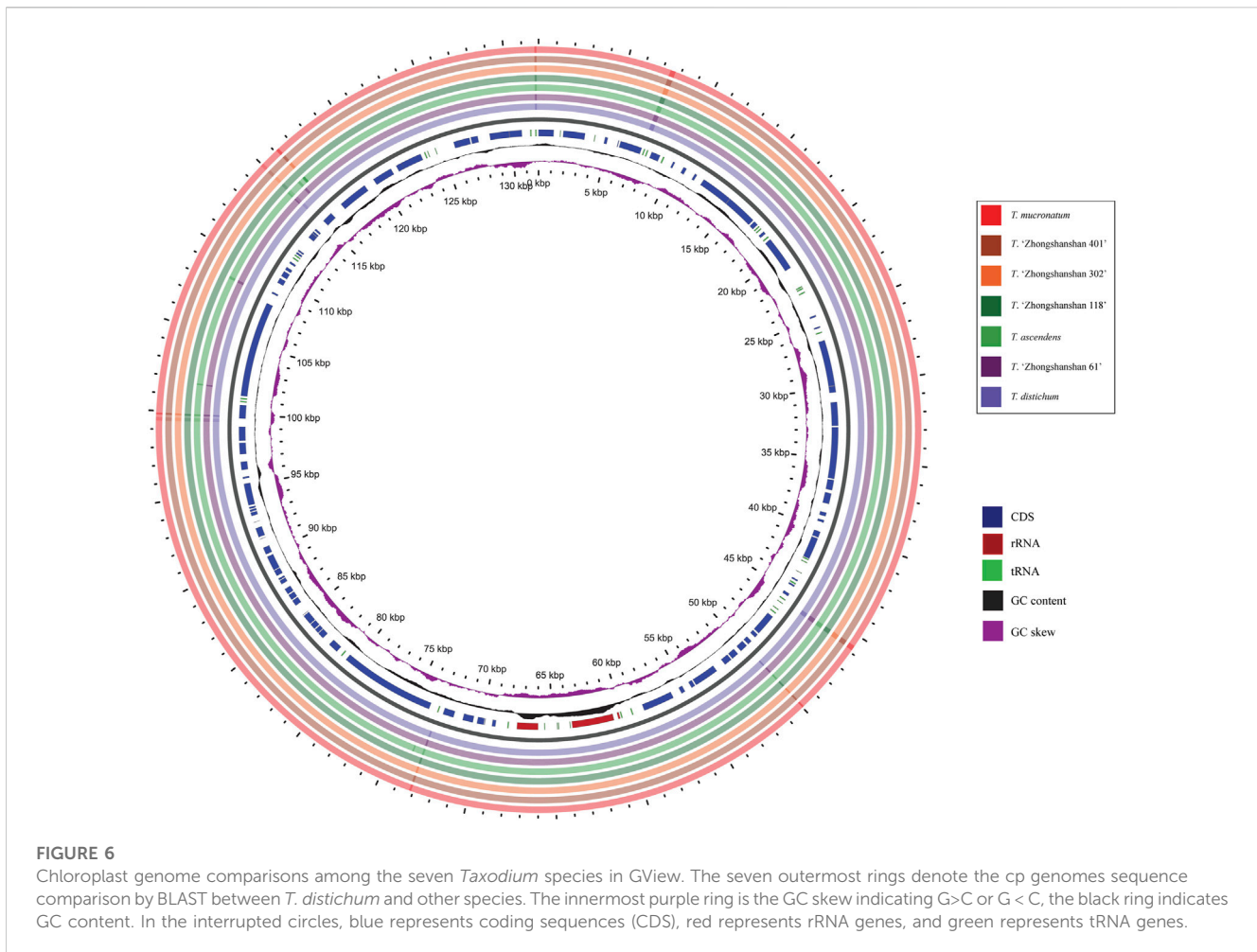
*ascendens*; 12 sequences of 30–50 bp and 17 sequences of greater than 91 bp were identified (Figure 3B; Supplementary Table S4).

Because of the high polymorphism, reproducibility, and abundance of cp SSRs (cpSSRs) in plant genomes, they are some of the most important molecular markers used in genetic analyses of plant populations, evolutionary studies, and breeding programs (Jansen et al., 2007). In our study, mononucleotide and dinucleotide SSRs were the most common cpSSRs in the cp genomes of *T.* “Zhongshanshan”, and they mainly consisted of A and T. This might be related to the high AT content of the cp genome, which is consistent with the results of studies of the cp genomes of other taxa (Duan et al., 2020; Chen et al., 2022). Repeat sequences are important sources of genomic variation and play key roles in mediating genomic rearrangements; they are commonly used for the development of genetic markers for phylogenetic and population studies (Tangphatsornruang et al., 2010). Of the four long repeat types, forward and palindrome repeats were the most abundant type of repeat in the cp genomes of *Taxodium*, which is consistent with the results of previous studies of the cp genomes of members of the genus *Cupressus* and genus *Juniperus* (Chen et al., 2022). Reverse and complementary repeats were present in *T. distichum* and *T. mucronatum* but absent in *T. ascendens*. Furthermore, reverse and complementary repeats were absent in *Cryptomeria duclouxiana*, *J. chinensis*, *J. gaussonii*, *J. pingii*, and *J. procumbens* (Chen et al., 2022). This indicates that there is substantial variation in the frequency of these two repeat types among genera in the family Cupressaceae.

### 3.3 Codon usage and GC content in the cp genomes of *T.* “Zhongshanshan”

The usage of synonymous encodes can prevent deleterious mutations, which can result in codon degeneration (Morton, 2003). However, the use of synonymous codons in organisms such as plants is biased (Liu and Xue, 2005), and the relative synonymous codon usage (RSCU) is an indicator of the magnitude of the codon usage bias (Sharp and Li, 1987). We measured the codon usage frequency and RSCU of the seven *Taxodium* species. The total number of codons for protein-coding genes ranged from 24,739 to 24,823, and the patterns of codon usage in *T. mucronatum*, *T.* “Zhongshanshan 118”, *T.* “Zhongshanshan 302”, and *T.* “Zhongshanshan 401” were similar (Figure 4; Supplementary Table S5). Excluding the stop codons, leucine (2,677–2,694 codons, 10.82%–10.85%) was the most abundant amino acid, and cysteine (280 codons, 0.01%) was the least abundant amino acid, respectively, excluding the stop codons. (Figure 4; Supplementary Table S5). Synonymous bias was observed when RSCU A total of 64 codons encoding 21 amino acids were identified. Based on RSCU values, synonymous codon preference can be divided into: no preference (RSCU ≤ 1) and preference (RSCU > 1). In this study, 33 codons showed preference and 31 codons showed no preference for these species. Moreover, except for UUG, most of the preferred codons ended with A or U (Supplementary Table S5).

The GC content is one of the most important characteristics of cp genomes. Thus, the GC content of the first, second, third positions, and entire codons of 83 protein-coding genes (hereafter referred to as GC1, GC2, GC3, and GCall, respectively) was determined. GC1, GC2, GC3, and GCall showed significant differences between genes but almost no differences between genomes (except for *ycf1* and *ycf2*). Furthermore,



the GC content of the *ycf1* and *ycf2* genes (either GC1s, GC2s, GC3s, or GCcall) in *T. ascendens* and its offspring *T. "Zhongshanshan 61"* differed significantly from that of other species, indicating that these two genes could be used as markers to distinguish between them. Cluster analysis based on heat maps of GC1s, GC2s, GC3s, and GCcall revealed that the four offspring were more closely related to their paternal parents than their maternal parents (Figure 5).

### 3.4 Comparative analysis of the complete cp genomes of *Taxodium* species

BLAST analysis of the seven *Taxodium* cp genomes was conducted using the mVISTA program and GView software, with *Taxodium distichum* as the reference. The entire cp genome was highly conserved among the examined species, especially the coding regions (Figure 6). Furthermore, divergence in the non-coding regions among species was greater than divergence among coding regions (Figure 7). A high level of divergence was observed for *rps16-chlB*, *trnC-proB*, *proC1*, *atpI-atpH*, *ndhF-trnN*, *rrn16-trnV*, and *trnV-rps12*, which could be suitable candidates for the identification of *Taxodium* species. Additionally, DnaSP software was used to identify hotspots in the cp genomes to characterize patterns of nucleotide variability. Five

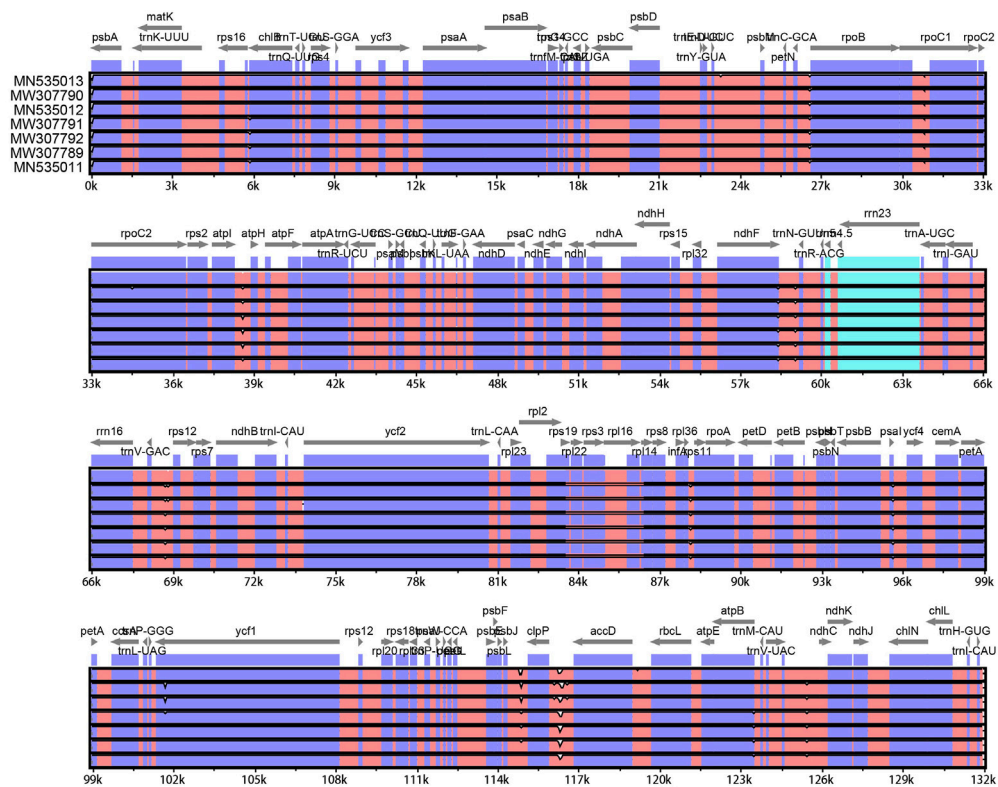
regions (*trnD-psbM*, *atpI-atpH*, *psbK-trnL*, *trnI-rrn16*, and *atpB-ndhC*) with *Pi* values were identified (>0.002), and all of these were located in non-coding regions (Supplementary Table S6; Supplementary Figure S1). Additional research is needed to determine whether these regions could be used as molecular markers in phylogenetic studies of *Taxodium* species.

Our findings indicated that the sequences of these seven cp genomes are highly conserved, especially in the coding regions, and the divergence in the non-coding regions among species is greater than the divergence in the coding regions. For example, high variation was observed among the *trnD-psbM*, *atpI-atpH*, *psbK-trnL*, *trnI-rrn16*, and *atpB-ndhC* regions. Research of these regions will aid future genetic studies as well as the development of molecular markers for the genus *Taxodium*.

### 3.5 Phylogenetic analysis

To characterize the evolutionary relationships among the four *Taxodium* cultivars obtained, MAFFT was used to align the entire cp genome sequences of the *Taxodium* species. The topology of the phylogenetic tree was visualized using IQ-tree software; the phylogenetic tree was built using the ML method with

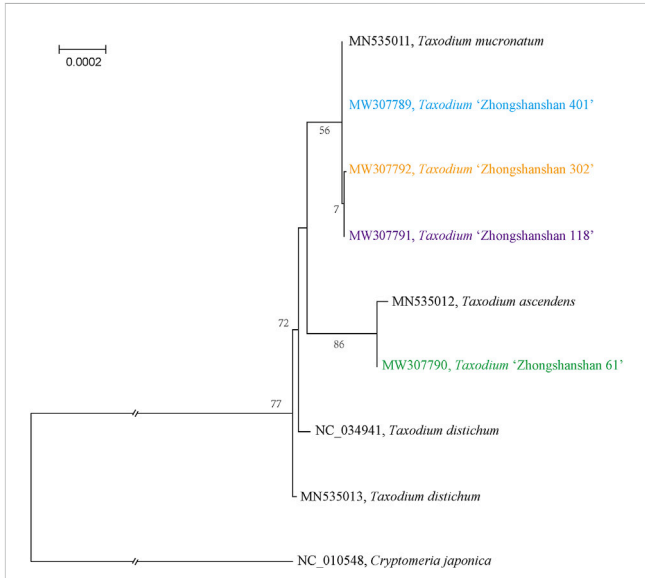




**FIGURE 7** A comparison of the cp genomes among seven *Taxodium* species. The vertical and horizontal axes in the figure represent the percentage of identity ranging from 50% to 100% and the sequence length, respectively. Annotated genes are displayed along the top.

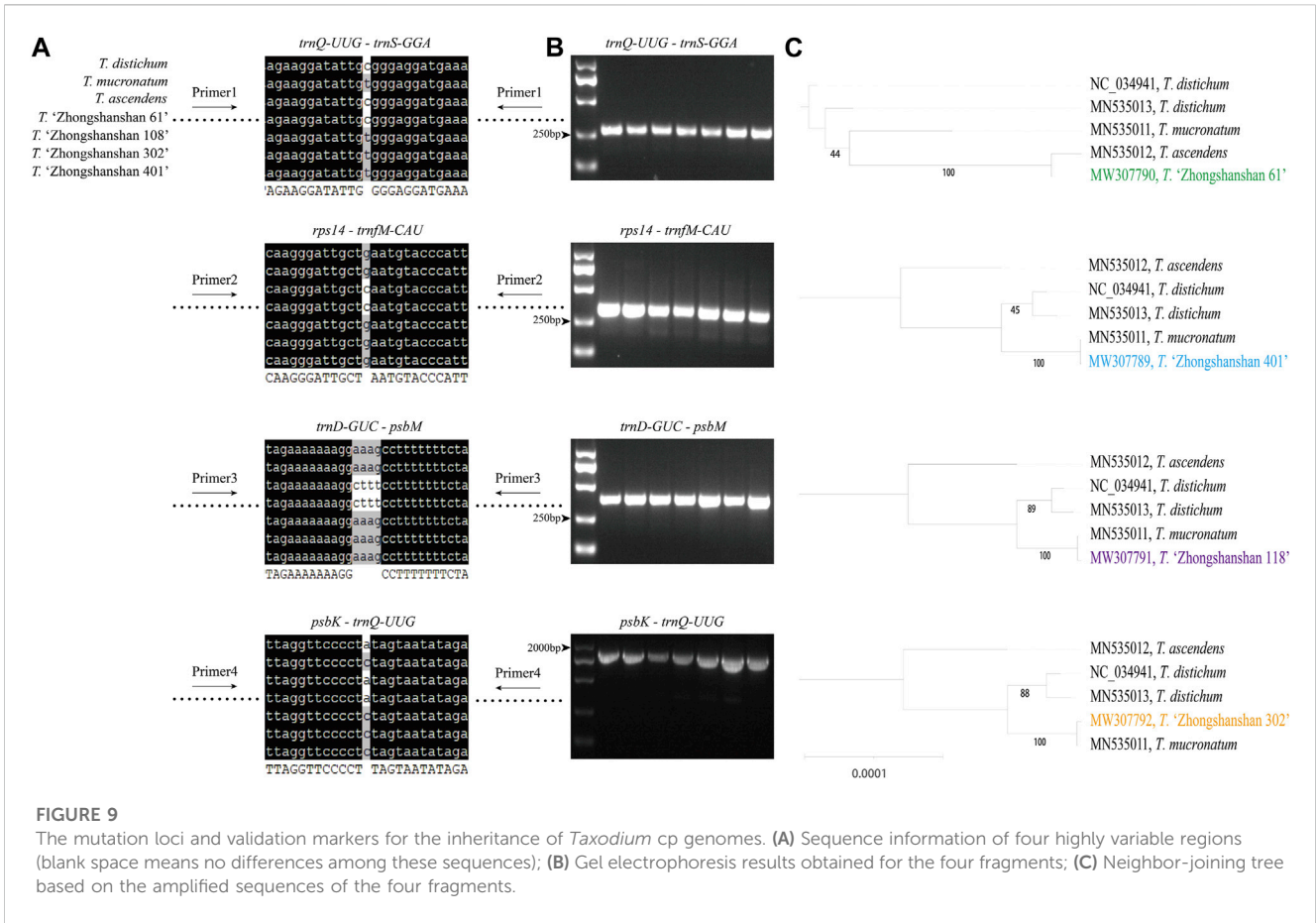
1,000 bootstrap replicates and with *Cryptomeria japonica* as the outgroup. *T.* “Zhongshanshan 61” was most closely related to *T. ascendens*, and *T.* “Zhongshanshan 401”, *T.* “Zhongshanshan 302”, *T.* “Zhongshanshan 118”, and *T. mucronatum* were closely related and comprised a subclade (Figure 8).

Two modes of inheritance of cp DNA (cpDNA) have been identified in seed plants cpDNA: uniparental (i.e., maternal or paternal) and biparental (i.e., maternal and paternal) (Hagemann, 1992; Crosby and Smith, 2012). The most common mode of cpDNA inheritance in seed plants is matrilineal, as only a few cases of patrilineal inheritance have been identified to date (Li et al., 2013). In gymnosperms, patrilineal inheritance is the most common mode of inheritance of cpDNA. Ohba (1971) was the first to report the paternal inheritance of cpDNA in reciprocal crosses of *C. japonica* (Ohba, 1971). Other species showing the paternal inheritance of cpDNA since then include *Pinus contora*, *Pinus banksiana* (Dong et al., 1992), *Pinus taeda* (Neale and Sederoff, 1989), and *Sequoia sempervirens* (Neale et al., 1989). The mechanisms underlying the paternal inheritance of cpDNA remain poorly studied. However, a mechanism underlying the paternal inheritance of cpDNA has been identified via studies of the formation of fertilized egg cells in *P. tabulaeformis*. During egg cell formation and development, the cpDNA of the oocyte is eliminated; consequently, no normal plasmid is present in the mature egg cells. At fertilization, the cytoplasm of the parent containing the cpDNA is transferred to the oocyte and then to the proto-embryo. The results of this study provide a cytological mechanism for the paternal

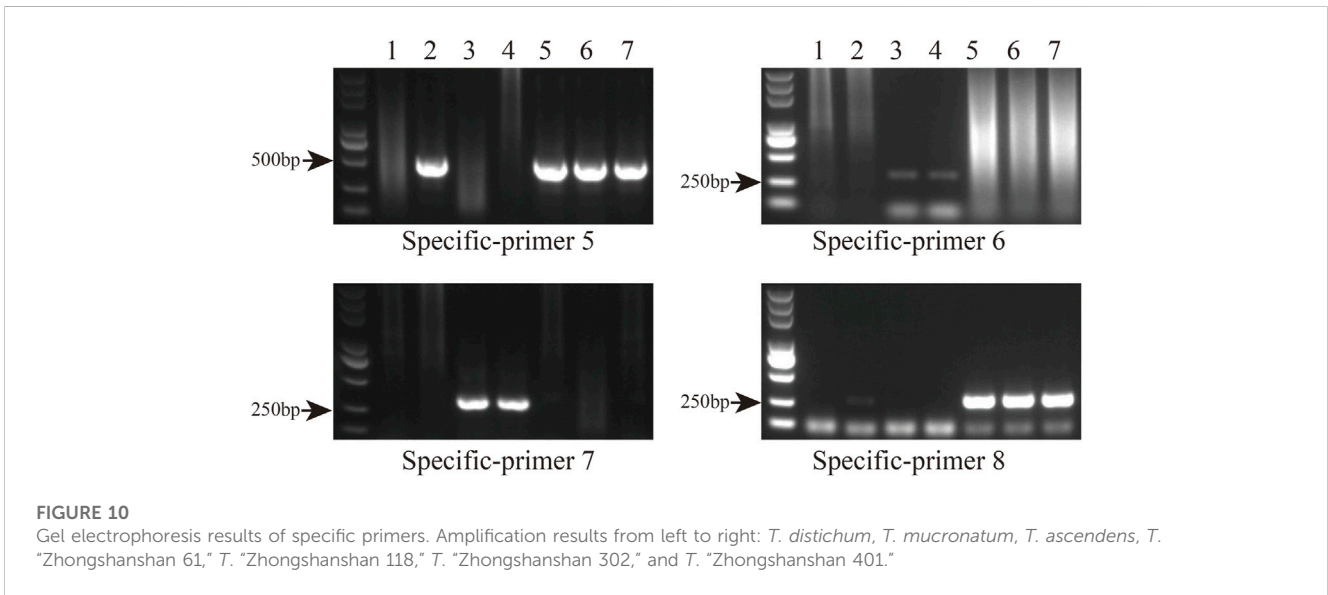


**FIGURE 8** Phylogenetic analysis of *Taxodium* species using their complete cp genomes. Numbers above the nodes are bootstrap values.

inheritance of cpDNA in *Pinus tabulaeformis*. Although the mechanism underlying the paternal inheritance of cpDNA in the genus *Taxodium* remains unclear, the results of this study have



**FIGURE 9** The mutation loci and validation markers for the inheritance of *Taxodium* cp genomes. (A) Sequence information of four highly variable regions (blank space means no differences among these sequences); (B) Gel electrophoresis results obtained for the four fragments; (C) Neighbor-joining tree based on the amplified sequences of the four fragments.



**FIGURE 10** Gel electrophoresis results of specific primers. Amplification results from left to right: *T. distichum*, *T. mucronatum*, *T. ascendens*, *T. 'Zhongshanshan 61'*, *T. 'Zhongshanshan 118'*, *T. 'Zhongshanshan 302'*, and *T. 'Zhongshanshan 401'*.

implications for genetic studies of *Taxodium* because of the major challenges associated with their large nuclear genomes. For example, the paternal parentage of hybrid offspring in open-pollinated seed orchards can be conveniently identified based on this trait of the cpDNA.

### 3.6 Molecular marker development

Primer sequences based on polymorphic loci in the cp genomes were designed using four fragments (Figure 9A; Supplementary Table S7). The cp genome of *Taxodium distichum* was used as

the reference sequence. First, a neighbor-joining (NJ) tree was built based on the amplified sequences of these four regions (Figure 9B). The results showed that the hybrid offspring clustered with their paternal parents (Figure 9C). In addition, single nucleotide polymorphism (SNP)/insertion and deletion (indel)-based specific primers were developed. Bands were observed for *T.* “Zhongshanshan 118”, *T.* “Zhongshanshan 302”, *T.* “Zhongshanshan 401”, and *T. mucronatum* when specific primers 5 and 8 were used for PCR amplification, and bands for *T. ascendens* and *T.* “Zhongshanshan 61” were only observed when specific primers 6 and 7 were used (Figure 10).

Complete cp genome information can provide insights into plant biology, diversity, and evolutionary relationships (Moore et al., 2007; Shaw et al., 2014). The sequence and structural characteristics of cp genomes make them effective DNA barcodes in plants. Hypervariable regions developed based on the cp genome information are becoming increasingly used for the identification of closely related plant species, such as *Ilex* (Chong et al., 2022), *Cyathula* (Guo et al., 2022), and *Solanum* (Kim and Park, 2019). The results of our study indicated that the cp genome of *Taxodium* is paternally inherited; the specific primers developed in our study could be used for the identification of the male parent of *Taxodium* hybrids. Although verification of these specific markers using other hybrids is needed, these markers could be useful in breeding and genetic research of *Taxodium* spp.

## 4 Conclusion

The complete cp genomes of four *T.* “Zhongshanshan” hybrids were sequenced and compared against the cp genomes of their parents. The full-length cp genomes of these four hybrids ranged from 131,942 to 132,128 bp, and all of the cp genomes contained 120 genes, including 83 protein-coding genes, 33 tRNAs, and four rRNAs. Highly variable regions (e.g., *rps16-chlB*, *trnC-proB*, *proC1*, *atpI-atpH*, *ndhF-trnN*, *rrn16-trnV*, and *trnV-rps12*) were detected, and these regions could aid future genetic and phylogenetic studies. Phylogenetic analysis of the evolutionary relationships based on the GC content of protein-coding genes indicated that the four hybrids were most closely related to their parents. In addition, molecular markers were developed using four hypervariable regions. Neighbor-joining trees were constructed based on PCR-amplified regions, as well as gel electrophoresis of specific primers, which were developed based on the SNPs/indels in these regions. Our findings suggest that these markers could be used to identify the male parent of the *Taxodium* hybrids. To our knowledge, this is the first study to demonstrate that the cp genomes of *Taxodium* are paternally inherited. The molecular markers developed in our study will aid *Taxodium* breeding research.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, MW307790 <https://www.ncbi.nlm.nih.gov/genbank/>, MW307791 <https://www.ncbi.nlm.nih.gov/genbank/>, MW307792

<https://www.ncbi.nlm.nih.gov/>, SRR13347061 <https://www.ncbi.nlm.nih.gov/>, SRR13347062 <https://www.ncbi.nlm.nih.gov/>, SRR13347063 <https://www.ncbi.nlm.nih.gov/>, SRR13347064.

## Author contributions

FZ designed the research study. MY and LX performed the research. CY, HC, and XL provided help and advice on sample collection. ML, XC, and YY analyzed the data. FZ and MY wrote the manuscript. FZ revised the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Natural Science Foundation of Jiangsu Province (grant no. BK20201244), the National Natural Science Foundation of China (grant no. 31971660), the National Natural Science Foundation of China (grant no. 32101488), and the Natural Science Foundation of Jiangsu Province (grant no. BK20210163).

## Acknowledgments

The authors sincerely thank Yanwei Zhou and Ting Zhou for the assistance provided with this study. In addition, the authors sincerely thank the reviewers for their careful reading and helpful comments on this manuscript.

## Conflict of interest

Author ML is employed by Guangzhou Bio&Data Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor J-YX declared a past co-authorship with the author CY.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1193023/full#supplementary-material>

## References

- Androsiuk, P., Jastrzębski, J. P., Pauksztó, L., Makowczenko, K., Okorski, A., Pszczółkowska, A., et al. (2020). Evolutionary dynamics of the chloroplast genome sequences of six *Colobanthus* species. *Sci. Rep.* 10 (1), 11522–11614. doi:10.1038/s41598-020-68563-5
- Chen, C., Xia, X., Peng, J., and Wang, D. (2022). Comparative analyses of six complete chloroplast genomes from the genus *Cupressus* and *Juniperus* (Cupressaceae). *Gene* 837, 146696. doi:10.1016/j.gene.2022.146696
- Cheng, Y., Yang, Y., Wang, Z., Qi, B., Yin, Y., and Li, H. (2015). Development and characterization of EST-SSR markers in *Taxodium* 'zhongshanshan'. *Plant Mol. Biol. Report.* 33, 1804–1814. doi:10.1007/s11105-015-0875-9
- Chong, X., Li, Y., Yan, M., Wang, Y., Li, M., Zhou, Y., et al. (2022). Comparative chloroplast genome analysis of 10 *Ilex* species and the development of species-specific identification markers. *Ind. Crop. Prod.* 187, 115408. doi:10.1016/j.indcrop.2022.115408
- Creech, D., Zhou, L., Yunlong, Y., and Eguluz-Piedra, T. (2011). Can *Taxodium* be improved. *Arnoldia* 69 (2), 11–20.
- Crosby, K., and Smith, D. R. (2012). Does the mode of plastid inheritance influence plastid genome architecture? *Plos one* 7 (9), 46260. doi:10.1371/journal.pone.0046260
- Denny, G. C., and Arnold, M. A. (2007). Taxonomy and nomenclature of baldcypress, pondcypress, and montezuma cypress: One, two, or three species? *HortTechnology* 17 (1), 125–127. doi:10.21273/HORTTECH.17.1.125
- Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45 (4), e18–e19. doi:10.1093/nar/gkw955
- Dong, J., Wagner, D., Yanchuk, A., Carlson, M., Magnussen, S., Wang, X.-R., et al. (1992). Paternal chloroplast DNA inheritance in *Pinus consora* and *Pinus banksiana*: Independence of parental species or cross direction. *J. Hered.* 83 (6), 419–422. doi:10.1093/oxfordjournals.jhered.a111244
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Available at: <https://worldveg.tind.io/record/33886/>.
- Duan, H., Guo, J., Xuan, L., Wang, Z., Li, M., Yin, Y., et al. (2020). Comparative chloroplast genomics of the genus *Taxodium*. *BMC Genom* 21 (1), 114. doi:10.1186/s12864-020-6532-1
- Guo, H., Wang, L., Xu, W., Huo, Z., Yang, P., Zhang, Q., et al. (2022). The complete chloroplast genome sequence of *Cyathula officinalis* and comparative analysis with four related species. *Gene* 839, 146728. doi:10.1016/j.gene.2022.146728
- Hagemann, R. (1992). "Plastid genetics in higher plants," in *Cell organelles* (Vienna: Springer).
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. U. S. A.* 104 (49), 19369–19374. doi:10.1073/pnas.0709121104
- Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20 (4), 1160–1166. doi:10.1093/bib/bbx108
- Kim, S., Park, T.-H., Kim, J. M., Ho, M. J., Kim, H. T., Song, S. H., et al. (2019). PCR-based markers developed by comparison of complete chloroplast genome sequences discriminate *Solanum chacoense* from other *Solanum* species. *J. Plant Biotechnol.* 46 (2), 79–89. doi:10.1016/j.yrtph.2019.01.011
- Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29 (22), 4633–4642. doi:10.1093/nar/29.22.4633
- Li, D., Qi, X., Li, X., Li, L., Zhong, C., and Huang, H. (2013). Maternal inheritance of mitochondrial genomes and complex inheritance of chloroplast genomes in actinidia lind.: Evidences from interspecific crosses. *Mol. Genet. Genomics* 288, 101–110. doi:10.1007/s00438-012-0732-6
- Liu, Q., and Xue, Q. (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. *J. Genet.* 84 (1), 55–62. doi:10.1007/BF02715890
- Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52 (5–6), 267–274. doi:10.1007/s00294-007-0161-y
- Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., et al. (2000). Vista: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16 (11), 1046–1047. doi:10.1093/bioinformatics/16.11.1046
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., et al. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37 (5), 1530–1534. doi:10.1093/molbev/msaa015
- Moore, M. J., Bell, C. D., Soltis, P. S., and Soltis, D. E. (2007). Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc. Natl. Acad. Sci. U. S. A.* 104 (49), 19363–19368. doi:10.1073/pnas.0708072104
- Morton, B. R. (2003). The role of context-dependent mutations in generating compositional and codon usage bias in grass chloroplast DNA. *J. Mol. Evol.* 56 (5), 612–629. doi:10.1007/s00239-002-2430-1
- Mudunuri, S. B., and Nagarajaram, H. A. (2007). IMEx: Imperfect microsatellite extractor. *Bioinformatics* 23 (10), 1181–1187. doi:10.1093/bioinformatics/btm097
- Neale, D. B., Marshall, K. A., and Sederoff, R. R. (1989). Chloroplast and mitochondrial DNA are paternally inherited in *Sequoia sempervirens* D. Don Endl. *Proc. Natl. Acad. Sci. U. S. A.* 86 (23), 9347–9349. doi:10.1073/pnas.86.23.9347
- Neale, D., and Sederoff, R. (1989). Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theor. Appl. Genet.* 77 (2), 212–216. doi:10.1007/BF00266189
- Ohba, K. (1971). Paternal transmission of a plastid anomaly in some reciprocal crosses of Sugi, *Cryptomeria japonica*. *Silvae Genet.* 20, 101–107.
- Palmer, J. D. (1987). Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. *Am. Nat.* 130, S6–S29. doi:10.1086/284689
- Petkau, A., Stuart-Edwards, M., Stothard, P., and Van Domselaar, G. (2010). Interactive microbial genome visualization with GView. *Bioinformatics* 26 (24), 3125–3126. doi:10.1093/bioinformatics/btq588
- Ping, J., Hao, J., Li, J., Yang, Y., Su, Y., and Wang, T. (2022). Loss of the IR region in conifer plastomes: Changes in the selection pressure and substitution rate of protein-coding genes. *Ecol. Evol.* 12 (1), 84999–e8514. doi:10.1002/ece3.8499
- Rice, P., Longden, I., and Bleasby, A. (2000). Emboss: The European molecular biology open software suite. *Trends Genet.* 16 (6), 276–277. doi:10.1016/S0168-9525(00)02024-2
- Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33 (2), 686–689. doi:10.1093/nar/gki366
- Sharp, P. M., and Li, W.-H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15 (3), 1281–1295. doi:10.1093/nar/15.3.1281
- Shaw, J., Shafer, H. L., Leonard, O. R., Kovach, M. J., Schorr, M., and Morris, A. B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *Am. J. Bot.* 101 (11), 1987–2004. doi:10.3732/ajb.1400398
- Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., Uthapaisanwong, P., Yoocha, T., Jomchai, N., et al. (2010). The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: Structural organization and phylogenetic relationships. *DNA Res.* 17 (1), 11–22. doi:10.1093/dnares/dsp025
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* 45 (1), 6–11. doi:10.1093/nar/gkx391
- Wang, Z., Cheng, Y., Yin, Y., Yu, C., Yang, Y., Shi, Q., et al. (2016). Genetic linkage map construction and QTL mapping of seedling height, basal diameter and crown width of *Taxodium* 'Zhongshanshan 302' × *T. mucronatum*. *Springerplus* 5 (1), 936–1013. doi:10.1186/s40064-016-2617-3
- Wang, Z., Gu, C., Xuan, L., Hua, J., Shi, Q., Fan, W., et al. (2017). Identification of suitable reference genes in *Taxodium* 'Zhongshanshan' under abiotic stresses. *Trees* 31 (5), 1519–1530. doi:10.1007/s00468-017-1566-y
- Xuan, L., Hua, J., Zhang, F., Wang, Z., Pei, X., Yang, Y., et al. (2021). Identification and functional analysis of ThADH1 and ThADH4 genes involved in tolerance to waterlogging stress in *Taxodium* hybrid 'Zhongshanshan 406'. *Genes* 12 (2), 225. doi:10.3390/genes12020225
- Yang, Y., Xuan, L., Yu, C., Wang, Z., Xu, J., Fan, W., et al. (2018). High-density genetic map construction and quantitative trait loci identification for growth traits in (*Taxodium distichum* var. *distichum* × *T. mucronatum*) × *T. mucronatum*. *BMC Plant Biol.* 18, 263–314. doi:10.1186/s12870-018-1493-0
- Yi, X., Gao, L., Wang, B., Su, Y.-J., and Wang, T. (2013). The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): Evolutionary comparison of *cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biol. Evol.* 5 (4), 688–698. doi:10.1093/gbe/evt042
- Zhang, F., Xuan, L., Chen, H., Yu, C., Chong, X., Yin, Y., et al. (2021). Comparative transcriptome profiling of resistant and susceptible *Taxodium* trees in responding to the infection by *Pestalotiopsis maculans*. *Forests* 12 (8), 1090–1108. doi:10.3390/f12081090