




## RESEARCH ARTICLE

# Loss of the IR region in conifer plastomes: Changes in the selection pressure and substitution rate of protein-coding genes

Jingyao Ping<sup>1</sup>  | Jing Hao<sup>1</sup> | Jinye Li<sup>1</sup> | Yiqing Yang<sup>2</sup> | Yingjuan Su<sup>3,4</sup>  | Ting Wang<sup>1</sup> 

<sup>1</sup>College of Life Sciences, South China Agricultural University, Guangzhou, China

<sup>2</sup>College of Life Science and Technology, Central South University of Forestry and Technology, Changsha, China

<sup>3</sup>School of Life Sciences, Sun Yat-sen University, Guangzhou, China

<sup>4</sup>Research Institute of Sun Yat-sen University, Shenzhen, China

## Correspondence

Yingjuan Su, Xingang West Road, Guangzhou, Guangdong, IA 510275, China.

Email: suyj@mail.sysu.edu.cn

Ting Wang, Wushan Road, Guangzhou, Guangdong, IA 510640, China.

Email: tingwang@scau.edu.cn

## Funding information

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

## Abstract

Plastid genomes (plastomes) have a quadripartite structure, but some species have drastically reduced or lost inverted repeat (IR) regions. IR regions are important for genome stability and the evolution rate. In the evolutionary process of gymnosperms, the typical IRs of conifers were lost, possibly affecting the evolutionary rate and selection pressure of genomic protein-coding genes. In this study, we selected 78 gymnosperm species (51 genera, 13 families) for evolutionary analysis. The selection pressure analysis results showed that negative selection effects were detected in all 50 common genes. Among them, six genes in conifers had higher  $\omega$  values than non-conifers, and 12 genes had lower  $\omega$  values. The evolutionary rate analysis results showed that 9 of 50 common genes differed between conifers and non-conifers. It is more obvious that in non-conifers, the rates of *psbA* (*trst*, *trsv*, *ratio*, *dN*, *dS*, and  $\omega$ ) were 2.6- to 3.1-fold of conifers. In conifers, *trsv*, *ratio*, *dN*, *dS*, and  $\omega$  of *ycf2* were 1.2- to 3.6-fold of non-conifers. In addition, the evolution rate of *ycf2* in the IR was significantly reduced. *psbA* is undergoing dynamic change, with an abnormally high evolution rate as a small portion of it enters the IR region. Although conifers have lost the typical IR regions, we detected no change in the substitution rate or selection pressure of most protein-coding genes due to gene function, plant habitat, or newly acquired IRs.

## KEYWORDS

conifer, gymnosperm, inverted repeat region, rate heterogeneity, selection pressure

## JEL CLASSIFICATION

Genomics

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

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The plastid genomes (plastomes) of most land plants have a highly conserved quadripartite structure, consisting of a large single-copy (LSC) region and a small single-copy (SSC) region separated by a pair of inverted repeat (IR) regions (Wicke et al., 2011). Due to selective pressure on photosynthesis-related elements, plastids have highly conserved gene content and order (Ruhlman & Jansen, 2014). And due to their highly conservation, a large copy number, lack of recombination, and uniparental inheritance, plastomes have been used to evaluate phylogeographic relationships, phylogeographic histories, and evolutionary events (Barrett et al., 2016; Julian et al., 2017; Moore et al., 2006; Pacheco et al., 2019; Shaw et al., 2014; Wang et al., 2020).

Inverted repeat regions are important in replication initiation (Heinhorst & Cannon, 1993), genomic stability (Maréchal & Brisson, 2010; Palmer & Thompson, 1982), and gene conservation (Palmer & Thompson, 1982; Wolfe et al., 1987). IR changes alter the size of the plastid genome (Kwon et al., 2020). From green algae to angiosperms, IR regions typically contain at least four rRNAs and five tRNAs (Mower & Vickrey, 2018). However, the IR region, as a hot spot of structural rearrangement, often contracts, expands, or is lost. In related species, the boundary of the IR region changes little, resulting in the gain or loss of a small number of genes (Downie & Jansen, 2015; Li et al., 2016; Ping, Li, et al., 2021; Wicke et al., 2014). The IR region of *Pelargonium*, *Psilotum*, and Trochodendraceae has undergone large-scale expansion, gaining a large number of genes from the SC region (Chumley et al., 2006; Grewe et al., 2013; Sun et al., 2013). The IR region is absent in the plastomes of some plant lineages, such as *prasinophyceans* (Lemieux et al., 2014), *trebouxiophyceans* (Turmel et al., 2015), *streptophytes* (Lemieux et al., 2016), Ulvophyceae (Cai et al., 2017), and Euglenaceae (Karnkowska et al., 2018) among algae; Pinaceae, Cupressophytes (Li, Gao, et al., 2016; Wu, Lin, et al., 2011; Wu, Wang, et al., 2011), and Taxaceae (Zhang et al., 2014) among gymnosperms; and the putranjivoid clade of Malpighiales (Jin et al., 2020), *Tahina spectabilis* (Arecaceae) (Barrett et al., 2016), legumes (Palmer & Thompson, 1981, 1982), *Carnegieia gigantea* (Cactaceae) (Sanderson et al., 2015), and *Erodium* species (Geraniaceae) (Blazier et al., 2016; Guisinger et al., 2011; Ruhlman et al., 2017) among angiosperms.

The sequence substitution rate in the IR region differs from that in the SC region. In some legumes with missing IR regions, the synonymous substitution rate of genes entering the SC region from the IR region is similar to that of those already in the SC region (Perry & Wolfe, 2002). The IR region has a low substitution rate in *Cycas* (Wu & Chaw, 2015). Zhu et al. (2016) found that the synonymous substitution rate of genes in the SC region in vascular plants (angiosperms, gymnosperms, and ferns) is 3.7-fold that of genes in the IR region; after the transfer of genes between the SC and IR regions, the substitution rate becomes consistent. Some genes in the IR region have high synonymous substitution rates in some species (*Pelargonium*,

*Plantago*, and *Silene*). In ferns, *psbA*, *rps7*, *3'-rps12*, and *ycf2* showed a reduced substitution rate and increased GC content after entering the IR region (Li, Kuo, et al., 2016). Recently, Ping, Feng, et al. (2021), Ping, Li, et al. (2021) reported that the substitution rate of *3'-rps12* in the IR region is lower than that of *5'-rps12* in the SC region in ferns and gymnosperms.

The IR region has important effects on genome stability and the evolutionary rate (Maréchal & Brisson, 2010; Palmer & Thompson, 1982; Wolfe et al., 1987). During gymnosperm evolution, conifers lost the IR region. *Araucaria cunninghamii* (Araucariaceae) is native to southeastern coastal areas of Oceania and is the main afforestation tree species in tropical and subtropical regions. It is one of the top five park tree species worldwide and an important garden ornamental plant, and it is cultivated widely in China. *Callitropsis funebris* (Cupressaceae) is endemic to China and distributed widely; it grows rapidly and has a wide range of uses and strong adaptability. Its phylogenetic classification is a focus of research (Zheng & Fu, 1978). As these two species are important representatives of their genera, we sequenced their plastomes. In this study, we selected 78 gymnosperms and 2 Polypodiales (outgroup) and analyzed selection pressure and the evolutionary rate using the maximum likelihood method. We investigated whether the selection pressure and evolutionary rate of protein-coding genes in conifers differed according to the presence of the IR region; whether the selection pressure and evolutionary rate of genes that enter the IR region differ from those that enter the SC region; and the heterogeneity of gymnosperm evolutionary rates.

## 2 | MATERIALS AND METHODS

### 2.1 | Sequencing and sequence preparation

Fresh leaves of *Callitropsis funebris* (E113°35', N23°15') and *Araucaria cunninghamii* (E113°200', N23°90') were taken from the campus of South China Agricultural University. Specimens were stored in Herbarium of the College of Life Sciences, SCAU (specimen no.: PJY-NYS1910 and PJY-BM1910). Plastid genomic DNA was extracted using the DNASECure Plant Genomic DNA Extraction Kit (Tiangen) and double-ended sequencing was performed on Illumina HiSeq2500 platform. Clean reads filtered by Trimmomatic V0.32 (Bolger et al., 2014) were spliced and assembled in Velvet V1.2.03 (Zerbino & Birney, 2008). The genes were predicted by the DOGMA Program (Wyman et al., 2004) and the plastome was mapped using the online site OGDRAW v1.3 (<https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>) after gene annotation (Greiner et al., 2019). Finally, the sequence information was submitted on the BankIt Platform (<https://www.ncbi.nlm.nih.gov/WebSub/>), and Genbank accession no. is MT227812 and MT227813, respectively.

In addition, 76 gymnosperms and 2 Polypodiales (as outgroups) were downloaded from the NCBI database. A total of 78 gymnosperms cover 13 families and 51 genera (Table 1). Genious prime 2020

TABLE 1 The Information of sampled species

Order	Family	Species name	GenBank accession no.	Species name	GenBank accession no.	
Cycadales	Cycadaceae	<i>Cycas revoluta</i>	NC_020319	<i>Cycas szechuanensis</i>	NC_042668	
		<i>Cycas panzhihuaensis</i>	NC_031413	<i>Cycas taitungensis</i>	NC_009618	
	Boweniaceae	<i>Bowenia serrulata</i>	NC_026036			
	Zamiaceae	<i>Stangeria eriopus</i>	NC_026041	<i>Zamia furfuracea</i>	NC_026040	
		<i>Ceratozamia hildae</i>	NC_026037	<i>Encephalartos lehmannii</i>	NC_027514	
		<i>Dioon spinulosum</i>	NC_027512	<i>Lepidozamia peroffskyana</i>	NC_027513	
		<i>Macrozamia mountperriensis</i>	NC_027511			
Ginkgoales	Ginkgoaceae	<i>Ginkgo biloba</i>	NC_016986			
Gnetales	Gnetaceae	<i>Gnetum montanum</i>	NC_021438	<i>Gnetum ula</i>	NC_028734	
		<i>Gnetum parvifolium</i>	NC_011942	<i>Gnetum gnemon</i>	NC_026301	
Ephedrales	Ephedraceae	<i>Ephedra equisetina</i>	NC_011954	<i>Ephedra intermedia</i>	NC_044772	
		<i>Ephedra foeminea</i>	NC_029347	<i>Ephedra sinica</i>	NC_044773	
Welwitschiales	Welwitschiaceae	<i>Welwitschia mirabilis</i>	EU342371			
Cupressales	Cupressaceae	<i>Cryptomeria japonica</i>	NC_010548	<i>Cupressus tonkinensis</i>	NC_039562	
		<i>Taiwania cryptomerioides</i>	NC_016065	<i>Cupressus gigantea</i>	NC_028155	
		<i>Taiwania flousiana</i>	NC_021441	<i>Cupressus sempervirens</i>	NC_026296	
		<i>Cunninghamia lanceolata</i>	NC_021437	<i>Callitropsis funebris</i>	MT227813	
		<i>Juniperus monosperma</i>	NC_024022	<i>Callitropsis nootkatensis</i>	KP099642	
		<i>Juniperus recurva</i>	NC_042763	<i>Callitropsis vietnamensis</i>	KX832629	
		<i>Taxodium distichum</i>	NC_034941	<i>Hesperocyparis lusitanica</i>	MH121051	
		<i>Taxodium mucronatum</i>	NC_045277	<i>Chamaecyparis formosensis</i>	NC_034943	
		<i>Calocedrus formosana</i>	NC_023121	<i>Chamaecyparis hodginsii</i>	NC_036996	
		<i>Glyptostrobus pensilis</i>	NC_031354	<i>Thuja occidentalis</i>	NC_042177	
		<i>Metasequoia glyptostroboides</i>	NC_027423	<i>Thuja sutchuenensis</i>	NC_042176	
		<i>Callitris rhomboidea</i>	NC_034940			
		Taxaceae	<i>Cephalotaxus sinensis</i>	MF977938	<i>Taxus fuana</i>	NC_038099
			<i>Cephalotaxus oliveri</i>	NC_021110	<i>Pseudotaxus chienii</i>	NC_041503
	<i>Amentotaxus argotaenia</i>		NC_027581	<i>Torreya fargesii</i>	NC_029398	
	<i>Amentotaxus formosana</i>		NC_024945			
	Sciadopityaceae	<i>Sciadopitys verticillata</i>	NC_029734			
Pinales	Pinaceae	<i>Cedrus deodara</i>	NC_014575	<i>Picea neoveitchii</i>	NC_043913	
		<i>Pinus massoniana</i>	MF564195	<i>Keteleeria davidiana</i>	NC_011930	
		<i>Pinus yunnanensis</i>	NC_043856	<i>Tsuga chinensis</i>	NC_030630	
		<i>Pseudolarix amabilis</i>	NC_030631	<i>Larix sibirica</i>	NC_036811	
		<i>Pseudotsuga sinensis</i>	NC_016064	<i>Larix decidua</i>	NC_016058	
		<i>Abies fargesii</i>	NC_042775	<i>Abies fanjingshanensis</i>	NC_042777	
Araucariales	Podocarpaceae	<i>Nageia nagi</i>	NC_023120	<i>Retrophyllum piresii</i>	NC_024827	
		<i>Podocarpus lambertii</i>	NC_023805	<i>Dacrydium elatum</i>	NC_045880	
		<i>Dacrycarpus imbricatus</i>	NC_034942	<i>Manoao colensoi</i>	NC_044893	
	Araucariaceae	<i>Agathis dammara</i>	NC_023119	<i>Araucaria heterophylla</i>	NC_026450	
		<i>Wollemia nobilis</i>	KP259800	<i>Araucaria araucana</i>	NC_045394	
		<i>Araucaria cunninghamii</i>	MT227812	<i>Araucaria bidwillii</i>	NC_045395	
		<i>Araucaria angustifolia</i>	NC_039155			
Polypodiales	Polypodiaceae	<i>Lepisorus clathratus</i>	NC_035739			
	Athyriaceae	<i>Athyrium anisopterum</i>	NC_035738			

(Kearse et al., 2012) software was used to extract 50 protein-coding genes. And performed sequence alignment and correction through the ClustalW (codons) module in MEGA X (Kumar et al., 2018).

## 2.2 | Analysis of phylogenetic relationships

We constructed the phylogenetic relationships based on the common gene dataset (the two Polypodiales as the outgroup). MEGA X and PAUP4.0 (Swofford, 2002) were used to construct the NJ (neighbor-joining) tree and MP (maximum-parsimony) tree, respectively. RaxmlGUI2 software built ML (maximum-likelihood) tree with GTRGAMMAI substitution model and 1,000 bootstrap (Stamatakis, 2014). BI (Bayesian inference) tree was built with MrBayes 3.2.6 software (Huelsenbeck & Ronquist, 2001). We reconstructed a background tree combined with the published phylogenetic relationship to analyze selection pressure and evolution rate, finally.

## 2.3 | Analysis of selection pressure

Codeml program in PAML4.9 was used to analyze the selection pressure (Yang, 2004). The M0 (one-ratio) model assumes that each branch has the same value of  $\omega$  under the Branch model. The Model 2 (two-ratio) assumes the foreground branch and the background branch have different  $\omega$  values. The likelihood ratio test of M0 and Model 2 can be used to detect the difference of selection pressure between the foreground branch and the background branch.

## 2.4 | Analysis of evolutionary rate

We used HyPhy 2.2.4 software (Pond et al., 2005) to calculate the evolution rate, which is based on the maximum likelihood method and in the context of the phylogenetic tree. The transition rate (*trst*), transversion rate (*trsv*), and *trsv/trst* (*ratio*) of each branch were calculated under the nucleotide HKY85 substitution model with the local parameter. Similarly, the synonymous (*dS*) and non-synonymous (*dN*) substitution rates and *dN/dS* ( $\omega$ ) of each branch were calculated under the codon MG94xHKY85x3\_4 substitution model with the *rate Het.* parameter and 4 *rate classes*. IBM SPSS v22.0 (IBM Corporation, 2014) software was used to perform the Mann-Whitney U test and Spearman's rank correlations test on related parameters.

# 3 | RESULTS

## 3.1 | Plastome characteristics

The 76 gymnosperms were divided into two types according to the presence of a typical IR region: 56 gymnosperms lacked a typical IR region (SC-56; conifer) and 22 gymnosperms had a typical IR

region (gnetophytes, cycads, and *Ginkgo biloba*; IR-22; non-conifer). For SC-56 (Appendix S1), the genome size ranged from 117,720 bp (*K. davidiana*) to 146,723 bp (*A. heterophylla*), with an average of 130,858 bp. The GC content ranged from 34.3% (*T. sutchuenensis*) to 39.1% (*C. davidiana*), with an average of 36.2%.

For IR-22 (Table 2), the genome size ranged from 109,518 bp (*Ephedra equisetina*) to 166,341 bp (*Macrozamia mountperriensis*), with an average of 142,888 bp. The GC content ranged from 36.6% to 40.1%, with an average of 38.7%. The size of the IR region ranged from 17,732 bp (*G. biloba*) to 26,137 bp (*Ceratozamia hildae*). Compared with other groups, the plastome of the gnetophytes was the smallest, and the proportion of SSC region was also the smallest (7.2–9.3%), but the proportion of the IR region was largest (16%–18.9%). The three groups all contained the protein-coding genes *rps7* and *3'-rps12* in the IR region, and *ndhB* was present in *G. biloba* and cycads. All plants, except *G. biloba*, possessed *ycf2*. Individually, *3'-psbA* was present in *Welwitschia mirabilis*, *Gnetum parvifolium*, and *Gnetum gnemon*, *rpl32* was detected in *W. mirabilis*, and *rps15* was detected in *E. equisetina* (Figure 1).

Rank sum tests indicated no significant differences in genome size ( $p = .202$ ) and significant differences in GC content ( $p < .01$ ) between IR-22 and SC-56.

## 3.2 | Phylogenetic analysis

The 50 protein-coding genes comprised 30 photosynthetic system genes, 15 genetic systems, and 5 other genes (Table 3). The constructed phylogenetic results (Appendix S2) showed that the phylogenetic relationship constructed by the NJ method was the clearest, and each group formed a monophyletic branch. Using the BI and ML methods, we found that all groups, except the Pinales, formed a monophyletic branch. In the MP tree, the relationship between Cupressales and Araucariales was unclear.

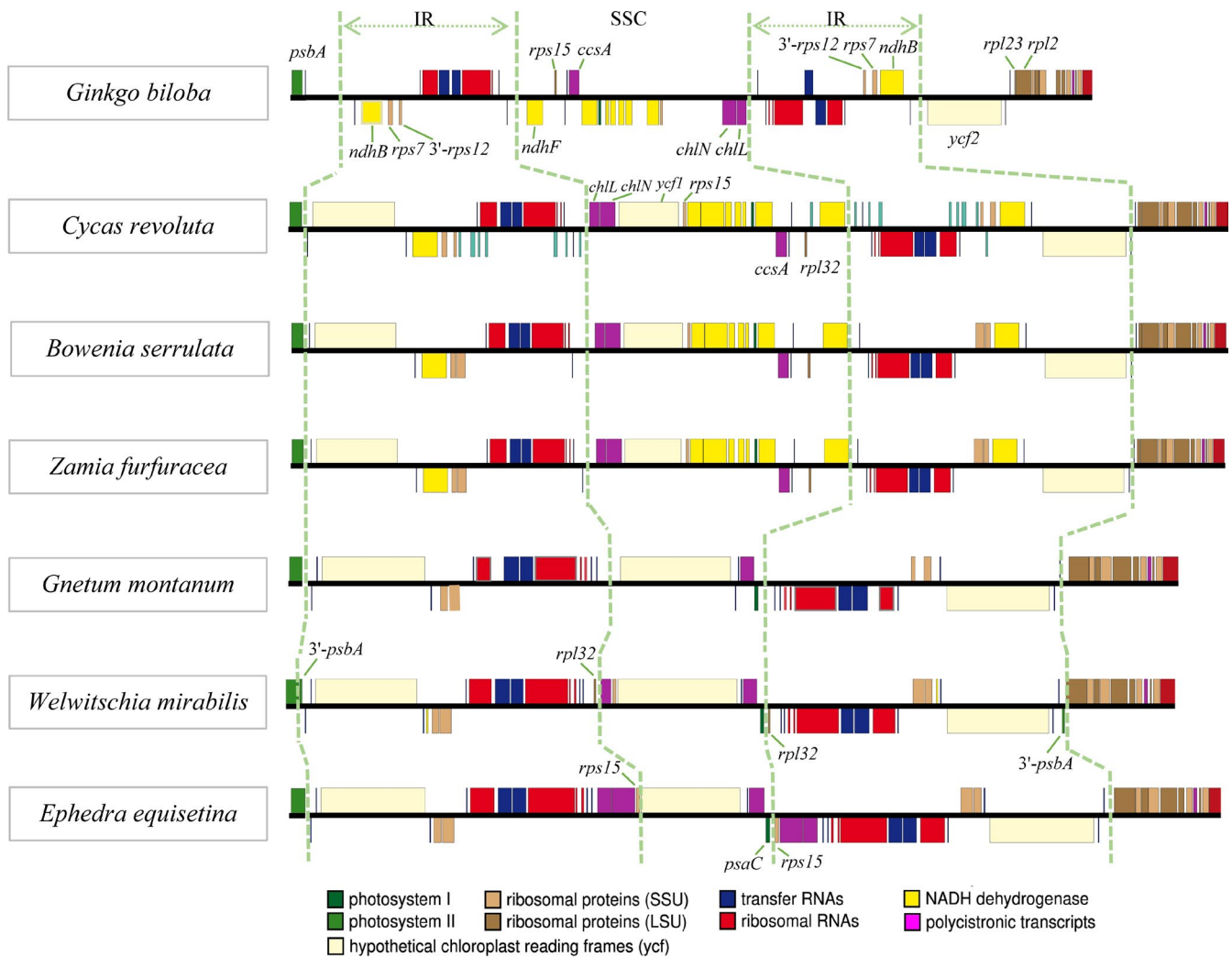
The NJ (bootstrap value = 100) and MP (bootstrap value = 50) trees supported gnetophytes as the basic group of gymnosperms, and the MP (bootstrap value = 100) and BI (posterior probability = 1) trees supported gnetophytes and *P. neoveitchii* as sister groups. Based on previous reports, we accepted the *Gnepine* hypothesis, and performed a manual adjustment to obtain the phylogenetic tree for selection pressure and evolution rate analysis (Figure 2).

## 3.3 | Evolutionary analysis based on the presence of a typical IR region

With IR-22 as the foreground branch and SC-56 as the background branch, likelihood ratio tests of M0 and Model 2 yielded 18 genes with significant differences ( $p < .05$ ). That is, these 18 genes experience different selection pressure in IR-22 and SC-56. Among them, six genes had an  $\omega$  value in IR-22 that were lower than SC-56, and the rest were higher (Table 4).

TABLE 2 Plastome structure information of 22 species with typical IR regions

Species name	Genome size/bp	GC/%	LSC/bp	LSC percentage	SSC/bp	SSC percentage	IR/bp	IR percentage	Number of genes in the IR region	
									CDS	tRNA
<i>Ginkgo biloba</i>	156,988	39.6	99,254	63.2%	22,267	14.2%	17,732	11.3%	3	6
<i>Cycas revoluta</i>	162,489	39.4	88,978	54.8%	23,379	14.4%	25,066	15.4%	4	6
<i>Cycas panzhihuaensis</i>	162,470	39.4	88,932	54.7%	23,448	14.4%	25,045	15.4%	4	7
<i>Cycas szechuanensis</i>	162,083	39.4	88,970	54.9%	23,107	14.3%	25,003	15.4%	4	7
<i>Cycas taitungensis</i>	163,403	39.5	90,216	55.2%	23,039	14.1%	25,074	15.3%	4	7
<i>Bowenia serrulata</i>	165,695	39.9	90,733	54.8%	23,156	14.0%	25,903	15.6%	4	7
<i>Stangeria eriopus</i>	163,548	39.5	89,850	54.9%	23,006	14.1%	25,346	15.5%	4	7
<i>Ceratozamia hildae</i>	165,734	39.7	90,487	54.6%	22,973	13.9%	26,137	15.8%	4	7
<i>Dioon spinulosum</i>	161,815	40.1	88,754	54.8%	23,355	14.4%	24,853	15.4%	4	7
<i>Zamia furfuracea</i>	164,953	39.7	90,441	54.8%	23,228	14.1%	25,642	15.5%	4	7
<i>Encephalartos lehmannii</i>	165,822	39.9	90,724	54.7%	23,302	14.1%	25,898	15.6%	4	7
<i>Lepidozamia peroffskyana</i>	165,939	39.9	90,804	54.7%	23,299	14.0%	25,981	15.7%	4	7
<i>Macrozamia mountperriensis</i>	166,341	39.8	91,171	54.8%	23,334	14.0%	25,918	15.6%	4	7
<i>Gnetum montanum</i>	115,019	38.2	66,139	57.5%	9,494	8.3%	19,693	17.1%	3	8
<i>Gnetum parvifolium</i>	114,914	38.2	66,095	57.5%	9,559	8.3%	19,630	17.1%	4	8
<i>Gnetum ula</i>	113,249	38.5	64,914	57.3%	8,791	7.8%	19,772	17.5%	3	8
<i>Gnetum gnemon</i>	115,022	38.2	66,591	57.9%	8,329	7.2%	20,051	17.4%	5	8
<i>Ephedra equisetina</i>	109,518	36.6	59,936	54.7%	8,078	7.4%	20,752	18.9%	6	8
<i>Ephedra foeminea</i>	109,584	36.7	60,708	55.4%	8,078	7.4%	20,399	18.6%	6	8
<i>Ephedra intermedia</i>	109,667	36.6	59,936	54.7%	8,247	7.5%	20,742	18.9%	6	8
<i>Ephedra sinica</i>	109,550	36.7	59,961	54.7%	8,103	7.4%	20,743	18.9%	6	8
<i>Welwitschia mirabilis</i>	119,726	36.7	68,556	57.3%	11,156	9.3%	20,007	16.7%	5	8



**FIGURE 1** Comparison of IR regions between different taxa. Different color blocks represent different gene types. In species with typical IR regions, there were two protein-coding genes (*rps7* and *3'-rps12*) located in the IR region. In *Ginkgo biloba*, the *ycf2* was located in the LSC region due to contraction of the IR region. Gnetophytes lost *ndh* genes, and in *Welwitschia mirabilis*, *3'-psbA* entered IR region

We analyzed and tested the evolutionary rates of the 50 selected common genes (Appendix S3). Among these, we found that the related parameters of nine genes were significantly different between IR-22 and SC-56 ( $p < .05$ ; Figure 3). Compared with SC-56, the genes with a higher evolution rates in IR-22 were *trst* (.003 vs. .001), *trsv* (.026 vs. .010), *ratio* (.082 vs. .031), *dN* (.007 vs. .001), *dS* (.070 vs. .025), and  $\omega$  (.100 vs. .032) of *psbA* (2.6- to 4.2-fold that of the SC-56); *trst* (.052 vs. .021), *dN* (.033 vs. .014), *dS* (.112 vs. .051), and  $\omega$  (.631 vs. .226) of *rps8* (2.3- to 2.4-fold); *dN* (.010 vs. .002) of *psbE* (5-fold); and  $\omega$  (.064 vs. .002) of *petG* (32-fold). Compared with IR-22, the genes with higher evolution rates in the SC-56 were *trsv* (.033 vs. .009), *ratio* (.504 vs. .145), *dN* (.063 vs. .022), *dS* (.068 vs. .024), and  $\omega$  (.904 vs. .024) of *ycf2* (1.2- to 3.6-fold those of IR-22); *trsv* (.008 vs. .005) of *rps4* (1.8-fold); and ratios of *rpoC1* (.224 vs. .087), *rbcL* (.234 vs. .077), and *cemA* (.183 vs. .064) (2.6-, 3-, and 2.8-fold, respectively). Spearman's rank correlation test results showed that the evolutionary rate of genes

other than *dS* of *rps8* (Spearman's rho = .157,  $p = .169$ ) and  $\omega$  of *psbA* (Spearman's rho = .212,  $p = .063$ ) was significantly correlated ( $p < .05$ ) with the IR region (Appendix S4).

The evolutionary rate of *psbA* (Appendix S5) was abnormally high in *Cycas taitungensis* and *G. biloba* and in *Cycas* spp. By contrast, the species evolutionary rate of SC-56 was low. For *rps8* (Appendix S6), evolutionary rates were high in *G. biloba*, *W. mirabilis*, and cycads; the values for most other species branches in SC-56 were 0, except *C. rhomboidea* and *S. verticillata*. For *rps4* (Appendix S7A), the transversion rate of IR-22 was low, and most values were 0. All species branches in SC-56 had high transversion rates, with those of *C. rhomboidea* and *S. verticillata* being abnormally high. For *psbE* (Appendix S7B), the *dN* of IR-22 was high (especially in *W. mirabilis* and *Dioon spinulosum*) and was 0 for most species in SC-56 ( $n = 49$ ). For *ycf2* (Appendix S8), substitution rates were high in most branches of SC-56, especially *C. rhomboidea* and *S. verticillata*. In IR-22, the value of the other branches was 0, except in *W. mirabilis* and *G. biloba*.



TABLE 3 Types of common genes

Gene type	Gene name
Genes for photosynthesis	
Photosystem	<i>psaA psaB psaC psal psaJ</i>
Photosystem II	<i>psbA psbB psbC psbD psbE psbF psbH psbl psbJ psbK psbL psbM psbN psbT</i>
Genetic system genes	
Cytochrome	<i>petA petB petD petG petL petN</i>
ATP Synthase	<i>atpB atpE atpF atpI</i>
RubiscoCO large subunit	<i>rbcl</i>
Ribosomal Proteins (LSU)	<i>rpl14 rpl20 rpl33 rpl36</i>
Ribosomal Proteins (SSU)	<i>rps2 rps4 rps7 rps8 rps11 rps18 rps19</i>
RNA Polymerase	<i>rpoA rpoB rpoC1 rpoC2</i>
Other genes	
Envelop membrane protein	<i>cemA</i>
C-type cytochrome synthesis	<i>ccsA</i>
Hypothetical chloroplast reading frames	<i>ycf2 ycf3 ycf4</i>

### 3.4 | Evolutionary analysis based on genes entering the IR region

The IR regions of 22 species contained the protein-coding genes 3'-*rps12* and *rps7* (Figure 1). In *W. mirabilis*, *G. parvifolium*, and *G. gne- mon*, part of the *psbA* (3'-*psbA*) entered the IR region; thus, it was divided into two categories (IR-3 and SC-75). In *G. biloba*, *ycf2* has been left from the IR region; thus, it was divided into two categories (IR-21 and SC-57).

Likelihood ratio test results revealed that the selection pressure differed among categories. For *psbA*,  $\omega_{IR-3}$  was 0.03474, with IR-3 as the foreground branch, and  $\omega_{SC-75}$  was 0.0001, with SC-75 as the background branch. For *ycf2*,  $\omega_{IR-21}$  was 0.76124, with IR-21 as the foreground branch, and  $\omega_{SC-57}$  was 0.67832, with SC-57 as the background branch. The *trsv*, *ratio*, and  $\omega$  values of *psbA* were significantly higher ( $P_{trsv} = .048$ ,  $P_{ratio} = .003$ , and  $P_{\omega} = .027$ ) in the IR region than in the SC region, and the evolutionary rate of *ycf2* was significantly lower ( $p < .05$ ) in the IR region than in the SC region (Figure 4).

## 4 | DISCUSSIONS

### 4.1 | Changes in genome size

Evolution has decreased the gymnosperm genome size. Conifers and non-conifers have genomes of similar sizes, which is related to the large compact genome of gnetophytes (115,022–109,518 bp). This compact genome may increase the gnetophyte survival rate

in harsh and competitive environments (Wu et al., 2009). After the removal of gnetophytes, a significant difference was detected between groups ( $p < .01$ ). The shrinkage of the conifer plastome is a result mainly of the loss of the IR region (IRb in Pinaceae and IRa in cupressophytes). In addition, the reduction in the cupressophyte plastome size may be caused by intergenic shrinkage resulting from the mutational burden (Wu & Chaw, 2014). Pinaceae and cupressophytes have lost their typical IRs during evolution, but have acquired one or more short, novel IRs, some of which also exhibit recombination to generate genomic structural diversity (Guo et al., 2014).

Compared with that of ginkgo, the IR region of other non-conifers has expanded. Although the IR region in gnetophytes is smaller than that in cycads, it accounts for a large proportion of the genome and contains more genes. Gnetophytes lost the *ndh* gene and non-coding sequence, resulting in a compact, more economical genome (Wu et al., 2009). In Pinaceae, the IR comprises only the full *trnI-CAU* gene and, in most species, a portion of the *psbA* gene (Lin et al., 2010; Wu, Lin, et al., 2011).

The GC content was increased in non-conifers, which may be related to a GC bias in gene conversion in the IR region. Indeed, the IR region of cycads contains more A/T to GC substitutions (Wu & Chaw, 2015).

### 4.2 | Phylogenetic location of gnetophytes

The four phylogenetic trees constructed using a tandem dataset of 50 common protein-coding genes were not consistent, and the application of different algorithms yielded different topological structures. Compared with those produced by NJ and MP, the phylogenetic relationships constructed by ML and BI are more accurate (Hall, 2005). The ML and BI trees showed that gnetophytes and *P. neoveitchi* were sister groups with a high degree of support. Although Pinaceae was a non-monophyletic group, it was closest to the gnetophytes, supporting the *Gnepine* hypothesis.

The phylogenetic position of the gnetophytes has been debated (Mathews, 2009; Palmer et al., 2004). Burleigh and Mathews (2004) examined 13 genes in gymnosperms (five plastid, four nuclear, and four mitochondrial) and almost 19 000 nucleotides, providing evidence for the *Gnepine* hypothesis. Wu and Chaw (2014) analyzed the rearrangement of DNA in the plastome of gymnosperms, and the results supported the *Gnepine* hypothesis. Li et al. (2017) analyzed single-copy genes from the genome and transcriptome, and the phylogenetic relationships constructed from their different datasets supported two topological structures: a sister relationship between gnetophytes and other gymnosperms and gnetophytes as a sister group to Pinaceae. Ran et al. (2018) constructed a phylogenetic tree with 1308 loci based on transcriptome data, which also supported the *Gnepine* hypothesis. Ping, Feng, et al. (2021) constructed ML and BI trees using *rbcl* and *matK*, which supported the classification of gnetophytes and Pinaceae as sister groups.

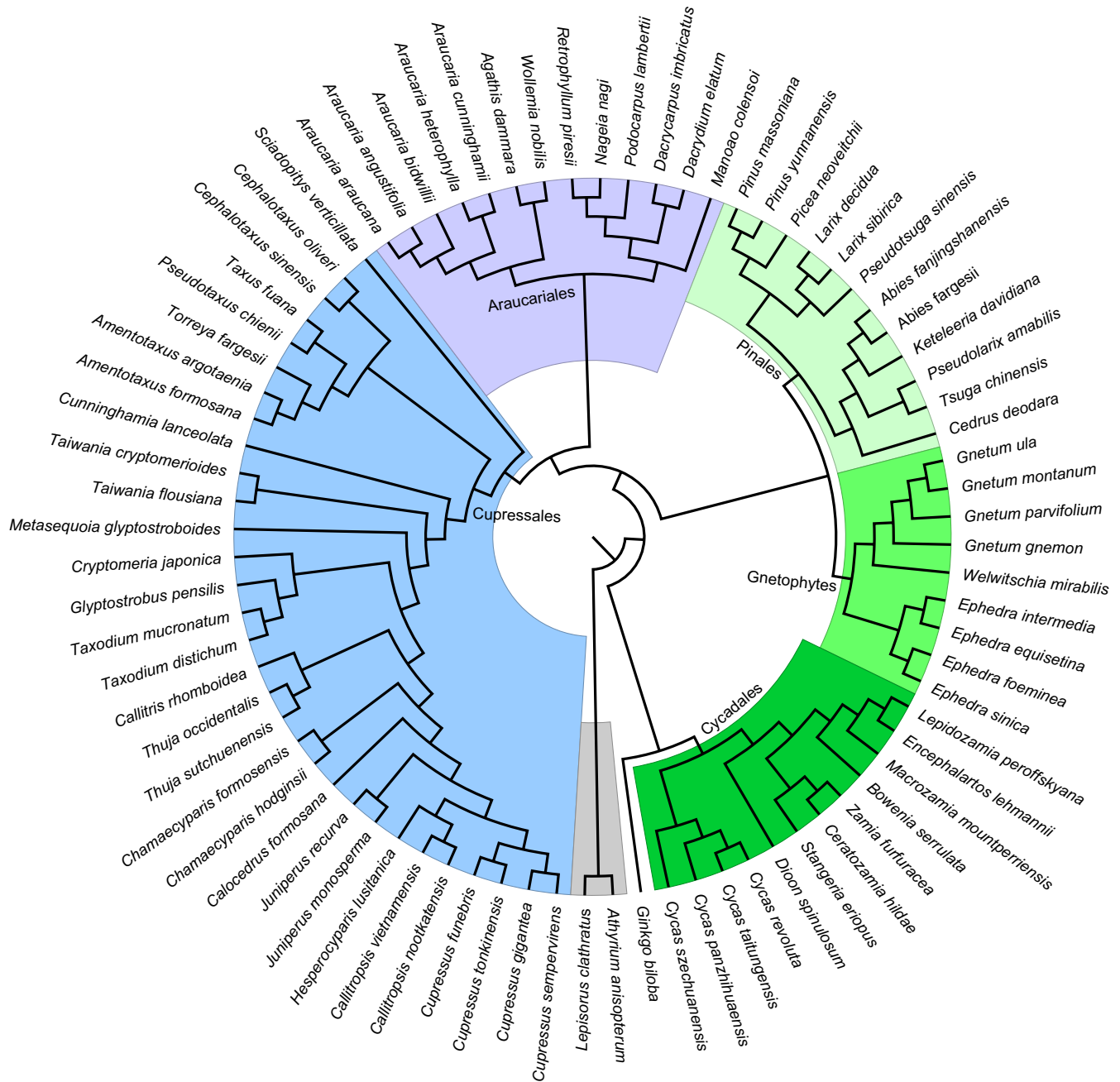


FIGURE 2 Phylogenetic relationship of sampled species. The tree was obtained through manual adjustment by combining four methods and previous research results. See Appendix S2 for four trees

### 4.3 | Heterogeneity of evolutionary rates

The change in evolutionary rate was independent of the presence of the IR region, but rather indicates an increased substitution rate of gnetophytes (especially *W. mirabilis*) and a lower evolutionary rate of conifers. The evolutionary rate of four genes was increased in non-conifers. The increased substitution rates of *rps8*, *psbE*, and *psbA* were related to the low evolutionary rate of conifers. The *petG* value was abnormally high. The *dN* or *dS* value for most species branches was 0, resulting in false-positive results. By contrast, the substitution rates of *rps4* and *ycf2* were decreased in non-conifers, which is related to

their generally low evolutionary rate. The evolutionary rate of most genes in *W. mirabilis* was high. The phylogenetic tree constructed using 51 common genes showed that the gnetophytes, especially *W. mirabilis* and *Ephedra sinica*, had high evolutionary rates. McCoy et al. (2008) reported that about 75% of plastid protein-coding genes showed high substitution rates in *W. mirabilis*. The substitution rates of most protein-coding genes did not differ significantly between conifers and non-conifers, which may be related to the newly acquired short IR region of conifers (Hirao et al., 2008; Wu & Chaw, 2014; Yi et al., 2013). In *Cephalotaxus oliveri*, the 544-bp IR (repeated *trnQ*-UUG gene) was inferred to have recombination activity (Yi et al., 2013).



TABLE 4 Genes that accepted Model 2 in the Branch Model

Gene	lnL M0	lnL Model 2	2 $\Delta$ l	P-value	$\omega_{\text{foreground}}$	$\omega_{\text{background}}$	$\omega_{\text{foreground}}/\omega_{\text{background}}$
<i>atpF</i>	-6580.533	-6587.056	13.046	0	0.367	0.195	1.882
<i>ccsA</i>	-11747.691	-11749.697	4.012	.045	0.268	0.346	<b>0.775</b>
<i>cemA</i>	-10261.516	-10266.192	9.352	.002	0.418	0.272	1.537
<i>petD</i>	-3345.317	-3348.100	5.566	.018	0.066	0.117	<b>0.564</b>
<i>psaA</i>	-16296.555	-16299.080	5.050	.025	0.078	0.057	1.368
<i>psaJ</i>	-1520.374	-1522.340	3.932	.047	0.269	0.132	2.038
<i>psbA</i>	-6124.730	-6133.245	17.029	0	0.026	0.071	<b>0.366</b>
<i>psbC</i>	-9768.612	-9773.699	10.173	.001	0.042	0.073	<b>0.575</b>
<i>psbE</i>	-1676.534	-1681.768	10.468	.001	0.058	0.183	<b>0.317</b>
<i>psbJ</i>	-1229.013	-1231.721	5.416	.020	0.133	0.366	<b>0.363</b>
<i>psbM</i>	-1381.099	-1384.893	7.588	.006	0.479	0.132	3.629
<i>rpoB</i>	-38294.175	-38327.819	67.289	0	0.286	0.156	1.833
<i>rpoC1</i>	-26817.747	-26840.736	45.978	0	0.357	0.191	1.869
<i>rpoC2</i>	-50272.622	-50279.519	13.795	0	0.372	0.294	1.265
<i>rps2</i>	-8506.349	-8508.885	5.071	.024	0.281	0.205	1.371
<i>rps11</i>	-4972.071	-4979.642	15.142	0	0.339	0.152	2.230
<i>rps18</i>	-5208.180	-5212.761	9.163	.002	0.352	0.160	2.200
<i>ycf2</i>	-142683.719	-142688.968	10.499	.001	0.768	0.635	1.209

Note: Foreground is IR-22 and background is SC-56; Bold font indicates that  $\omega_{\text{foreground}}$  is less than  $\omega_{\text{background}}$ .

Gnetophytes have a higher substitution rate than do other gymnosperms (Ran et al., 2018; Wang et al., 2015; Wu et al., 2009), possibly because of higher mutation rates, changes in selective pressure, increased fixation of mutations by genetic drift, and biological characteristics (Fry & Wernegreen, 2005; Lanfear et al., 2010). Wu et al. (2009) proposed that the high frequency of AT-rich codons in gnetophytes leads to a higher substitution rate. Wang et al. (2015) proposed that the generation time and plant height underlie the increased substitution rate of gnetophytes (Lanfear et al., 2010, 2013). Ran et al. (2018) found that gnetophytes and angiosperms have similar rates of molecular evolution, which are higher than those of other gymnosperms, suggesting that gnetophytes and angiosperms experienced similar selection pressure during their evolutionary histories. As conifers are typically taller than non-conifers, the long-term rate of mitosis in the apical meristem is slower, the frequency of DNA replication is lower, and the accumulated errors per unit time are reduced, resulting in low mutation and substitution rates (Lanfear et al., 2013).

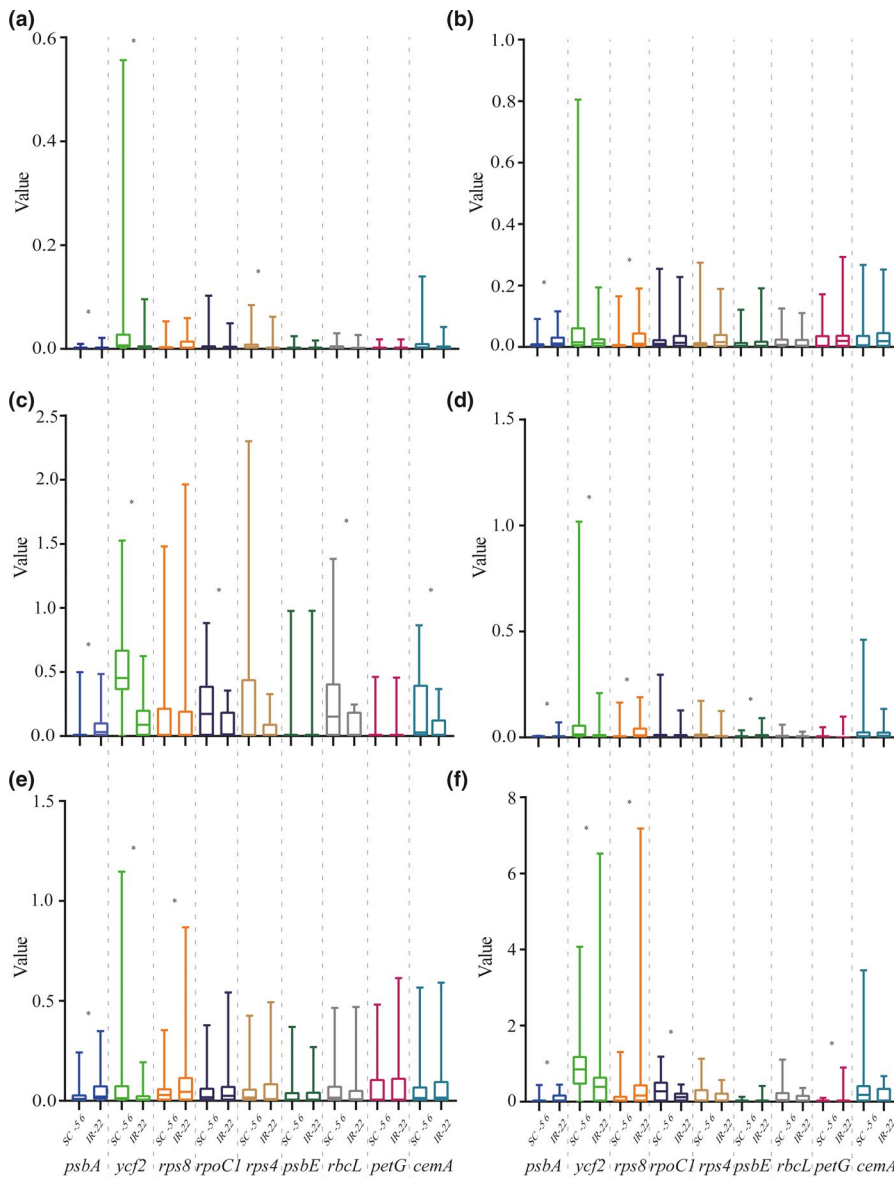
#### 4.4 | Negative selection pressure

Genes in the IR region experience strong negative selection (Ping, Feng, et al., 2021; Ping, Li, et al., 2021). The IR region is important for genome structural stability, and genomes lacking an IR region may experience different natural selection effects. However, the selection pressure of most genes did not differ significantly, indicating that they are not affected by the IR region. When  $\omega < 1$ , smaller  $\omega$  values reflect stronger negative selection pressure (Yang, 2006).

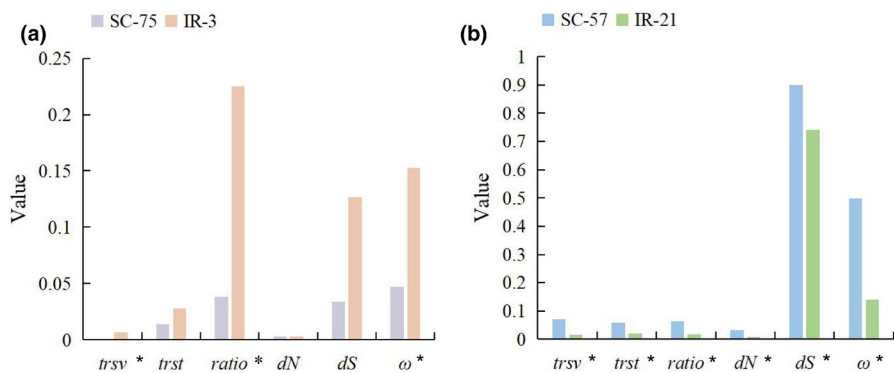
The selection pressure of 18 genes differed between conifers and non-conifers, and their  $\omega$  values were  $< 1$  (Table 4). Compared with conifers, six genes (four PSII genes, one cytochrome gene, and one c-type cytochrome synthesis gene) showed lower  $\omega$  values in non-conifers, and 12 genes (three PSI genes, one PSII gene, six genetic system genes, and two other genes) showed higher  $\omega$  values in non-conifers. The lower  $\omega$  values of these genes indicates the greater the negative selection effect, which may be more conducive to their function in various groups. The differences in selection pressure between conifers and non-conifers may be related to plant height; non-conifers are typically shorter than conifers. Efficient photosynthesis is required to obtain enough sunlight, necessitating negative selection pressure on the genes encoding the photosynthetic machinery. However, some such genes in conifers experienced negative selection pressure, which may be related to their structure or function or to habitat differences. The negative selection exerted by the habitats of gnetophytes leads to gene-specific reductions in  $dN/dS$  in their plastomes and genomes (Wang et al., 2015).

#### 4.5 | Reduced substitution rate in the IR region

Genes in the IR region tend to have a low substitution rate (Li, Kuo, et al., 2016; Li, Gao, et al., 2016; Ping, Feng, et al., 2021; Ping, Li, et al., 2021). In this study, *ycf2* in the IR region showed a reduced substitution rate (Figure 4) but weak selection pressure ( $\omega_{\text{IR-21}} = 0.76124$ ,  $\omega_{\text{SC-57}} = 0.67832$ ), indicating the lack of correlation between these factors. However, Lin et al. (2012) showed that the substitution



**FIGURE 3** Nine of fifty common genes with significantly different evolutionary rates between IR-22 (non-conifers) and SC-56 (conifers). (a) Transversion rate. (b) Transition rate. (c) Ratio. (d) Non-synonymous substitution rate. (e) Synonymous substitution rate. (f)  $\omega$ . \*: Rank sum test,  $p < .05$ . Compared with SC-56, the genes with a higher evolution rate in IR-22 were *trst*, *trsv*, *ratio*, *dN*, *dS*, and  $\omega$  of *psbA*; *trst*, *dN*, *dS*, and  $\omega$  of *rps8*; *dN* of *psbE*; and  $\omega$  of *petG*. Compared with IR-22, the genes with higher evolution rates in the SC-56 were *trsv*, *ratio*, *dN*, *dS*, and  $\omega$  of *ycf2*; *trsv* of *rps4*; and ratios of *rpoC1*, *rbcl*, and *cema*



**FIGURE 4** The evolutionary rate of *psbA* and *ycf2* in the different categories. (a) For *psbA*, the mean evolutionary rate of IR-3 and SC-75. B: For *ycf2*, the mean evolutionary rate of IR-21 and SC-57. \*: Rank sum test,  $p < .05$ . The *trsv*, *ratio*, and  $\omega$  values of *psbA* were significantly higher in the IR region than in the SC region, and the evolutionary rate of *ycf2* was significantly lower in the IR region than in the SC region

rate of *ycf2* in the SC region of *G. biloba* was not significantly higher than that in the IR region in seven other species. These discrepancies in results are attributable to differences in sample size. In gymnosperms, 3'-*rps12* (exons 2–3) shows a reduced substitution rate and negative selection pressure when present in the IR region (Ping, Feng, et al., 2021).

The substitution rate and selection pressure of *rps7* in the IR region did not change, indicating high stability during gymnosperm evolution. The other gene, *psbA*, encodes the D1 protein in the PSII core complex. In ferns, *psbA*, *ycf2*, *rps7*, and 3'-*rps12* (exons 2–3) showed reduced substitution rates when present in the IR region (Li, Kuo, et al., 2016; Li, Gao, et al., 2016). *psbA* only partially enters the

IR region in three gymnosperm species and shows increased substitution rates (Figure 4) and weak selection pressure. These characteristics differ from those of other genes; the most reasonable explanation is that this gene is undergoing dynamic changes in gymnosperms, resulting in changes in its substitution rate and selection pressure.

IR regions may play an important role in the maintenance of genomic stability (Maréchal & Brisson, 2010), such as by replication initiation (Heinhorst & Cannon, 1993), genome stabilization (Palmer & Thompson, 1982), and gene conservation (Palmer & Thompson, 1982; Wolfe et al., 1987). IR loss is related to structural recombination. In some taxa lacking an IR region, plastids underwent significant structural changes, including gene and intron loss, multiple inversion, and translocation and duplication of plastome segments (Jin et al., 2020; Hirao et al., 2008; Mower & Vickrey, 2018). Sequences in the IR region often have a reduced substitution rate, which is related to the region's characteristics. Due to its double-copy nature, the frequency of gene conversion is higher, and some mutation sites are repaired, resulting in a low substitution rate (Birky & Walsh, 1992; Khakhlova & Bock, 2006).

#### ACKNOWLEDGEMENTS

We thank the National Natural Science Foundation for its support (31670200, 31770587, 31872670, and 32071781).

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### AUTHOR CONTRIBUTION

**Jingyao Ping:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal). **Jing Hao:** Data curation (equal); Formal analysis (equal); Methodology (equal). **Jinye Li:** Data curation (equal); Formal analysis (equal); Resources (equal). **Yiqing Yang:** Formal analysis (equal); Methodology (equal). **Yingjuan Su:** Funding acquisition (equal); Writing – original draft (equal); Writing – review & editing (equal). **Ting Wang:** Funding acquisition (equal); Writing – original draft (equal); Writing – review & editing (equal).

#### DATA AVAILABILITY STATEMENT

Data source is NCBI database:

[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_020319](https://www.ncbi.nlm.nih.gov/nuccore/NC_020319)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_031413](https://www.ncbi.nlm.nih.gov/nuccore/NC_031413)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_042668](https://www.ncbi.nlm.nih.gov/nuccore/NC_042668)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_009618](https://www.ncbi.nlm.nih.gov/nuccore/NC_009618)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_026036](https://www.ncbi.nlm.nih.gov/nuccore/NC_026036)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_026041](https://www.ncbi.nlm.nih.gov/nuccore/NC_026041)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_026037](https://www.ncbi.nlm.nih.gov/nuccore/NC_026037)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_027512](https://www.ncbi.nlm.nih.gov/nuccore/NC_027512)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_026040](https://www.ncbi.nlm.nih.gov/nuccore/NC_026040)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_027514](https://www.ncbi.nlm.nih.gov/nuccore/NC_027514)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_027513](https://www.ncbi.nlm.nih.gov/nuccore/NC_027513)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_027511](https://www.ncbi.nlm.nih.gov/nuccore/NC_027511)

[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_016986](https://www.ncbi.nlm.nih.gov/nuccore/NC_016986)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_021438](https://www.ncbi.nlm.nih.gov/nuccore/NC_021438)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_011942](https://www.ncbi.nlm.nih.gov/nuccore/NC_011942)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_028734](https://www.ncbi.nlm.nih.gov/nuccore/NC_028734)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_026301](https://www.ncbi.nlm.nih.gov/nuccore/NC_026301)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_011954](https://www.ncbi.nlm.nih.gov/nuccore/NC_011954)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_029347](https://www.ncbi.nlm.nih.gov/nuccore/NC_029347)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_044772](https://www.ncbi.nlm.nih.gov/nuccore/NC_044772)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_044773](https://www.ncbi.nlm.nih.gov/nuccore/NC_044773)  
<https://www.ncbi.nlm.nih.gov/nuccore/EU342371>  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_010548](https://www.ncbi.nlm.nih.gov/nuccore/NC_010548)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_016065](https://www.ncbi.nlm.nih.gov/nuccore/NC_016065)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_021441](https://www.ncbi.nlm.nih.gov/nuccore/NC_021441)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_021437](https://www.ncbi.nlm.nih.gov/nuccore/NC_021437)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_024022](https://www.ncbi.nlm.nih.gov/nuccore/NC_024022)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_042763](https://www.ncbi.nlm.nih.gov/nuccore/NC_042763)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_034941](https://www.ncbi.nlm.nih.gov/nuccore/NC_034941)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_045277](https://www.ncbi.nlm.nih.gov/nuccore/NC_045277)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_023121](https://www.ncbi.nlm.nih.gov/nuccore/NC_023121)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_039562](https://www.ncbi.nlm.nih.gov/nuccore/NC_039562)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_028155](https://www.ncbi.nlm.nih.gov/nuccore/NC_028155)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_026296](https://www.ncbi.nlm.nih.gov/nuccore/NC_026296)  
<https://www.ncbi.nlm.nih.gov/nuccore/MT227813>  
<https://www.ncbi.nlm.nih.gov/nuccore/KP099642>  
<https://www.ncbi.nlm.nih.gov/nuccore/KX832629>  
<https://www.ncbi.nlm.nih.gov/nuccore/MH121051>  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_034943](https://www.ncbi.nlm.nih.gov/nuccore/NC_034943)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_036996](https://www.ncbi.nlm.nih.gov/nuccore/NC_036996)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_031354](https://www.ncbi.nlm.nih.gov/nuccore/NC_031354)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_027423](https://www.ncbi.nlm.nih.gov/nuccore/NC_027423)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_042176](https://www.ncbi.nlm.nih.gov/nuccore/NC_042176)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_042177](https://www.ncbi.nlm.nih.gov/nuccore/NC_042177)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_034940](https://www.ncbi.nlm.nih.gov/nuccore/NC_034940)  
<https://www.ncbi.nlm.nih.gov/nuccore/MF977938>  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_021110](https://www.ncbi.nlm.nih.gov/nuccore/NC_021110)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_027581](https://www.ncbi.nlm.nih.gov/nuccore/NC_027581)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_024945](https://www.ncbi.nlm.nih.gov/nuccore/NC_024945)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_038099](https://www.ncbi.nlm.nih.gov/nuccore/NC_038099)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_041503](https://www.ncbi.nlm.nih.gov/nuccore/NC_041503)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_029398](https://www.ncbi.nlm.nih.gov/nuccore/NC_029398)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_029734](https://www.ncbi.nlm.nih.gov/nuccore/NC_029734)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_014575](https://www.ncbi.nlm.nih.gov/nuccore/NC_014575)  
<https://www.ncbi.nlm.nih.gov/nuccore/MF564195>  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_043856](https://www.ncbi.nlm.nih.gov/nuccore/NC_043856)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_030631](https://www.ncbi.nlm.nih.gov/nuccore/NC_030631)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_016064](https://www.ncbi.nlm.nih.gov/nuccore/NC_016064)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_043913](https://www.ncbi.nlm.nih.gov/nuccore/NC_043913)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_011930](https://www.ncbi.nlm.nih.gov/nuccore/NC_011930)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_030630](https://www.ncbi.nlm.nih.gov/nuccore/NC_030630)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_036811](https://www.ncbi.nlm.nih.gov/nuccore/NC_036811)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_016058](https://www.ncbi.nlm.nih.gov/nuccore/NC_016058)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_042775](https://www.ncbi.nlm.nih.gov/nuccore/NC_042775)  
[https://www.ncbi.nlm.nih.gov/nuccore/NC\\_042777](https://www.ncbi.nlm.nih.gov/nuccore/NC_042777)

https://www.ncbi.nlm.nih.gov/nuccore/NC\_023120  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_023805  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_034942  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_024827  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_045880  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_044893  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_023119  
 https://www.ncbi.nlm.nih.gov/nuccore/KP259800  
 https://www.ncbi.nlm.nih.gov/nuccore/MT227812  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_039155  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_026450  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_045394  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_045395  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_035739  
 https://www.ncbi.nlm.nih.gov/nuccore/NC\_035738

## ORCID

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

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

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

## REFERENCES

- Barrett, C. F., Baker, W. J., Comer, J. R., Conran, J. G., Lahmeyer, S. C., Leebens-Mack, J. H., Li, J., Lim, G. S., Mayfield-Jones, D. R., Perez, L., Medina, J., Pires, J. C., Santos, C., Wm. Stevenson, D., Zomlefer, W. B., & Davis, J. I. (2016). Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. *New Phytologist*, 209(2), 855–870. <https://doi.org/10.1111/nph.13617>
- Birky, C. W. Jr, & Walsh, J. B. (1992). Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. *Genetics*, 130(3), 677–683. <https://doi.org/10.1093/genetics/130.3.677>
- Blazier, J. C., Jansen, R. K., Mower, J. P., Govindu, M., Zhang, J., Weng, M.-L., & Ruhlman, T. A. (2016). Variable presence of the inverted repeat and plastome stability in *Erodium*. *Annals of Botany*, 117(7), 1209–1220. <https://doi.org/10.1093/aob/mcw065>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Burleigh, J. G., & Mathews, S. (2004). Phylogenetic signal in nucleotide data from seed plants: Implications for resolving the seed plant tree of life. *American Journal of Botany*, 91(10), 1599–1613. <https://doi.org/10.3732/ajb.91.10.1599>
- Cai, C., Wang, L., Zhou, L., & Jiao, B. (2017). Complete chloroplast genome of green tide algae *Ulva flexuosa* (Ulvophyceae, Chlorophyta) with comparative analysis. *PLoS One*, 12(9), e0184196. <https://doi.org/10.1371/journal.pone.0184196>
- Chumley, T. W., Palmer, J. D., Mower, J. P., Fourcade, H. M., Calie, P. J., Boore, J. L., & Jansen, R. K. (2006). The complete chloroplast genome sequence of *Pelargonium x hortorum*: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Molecular Biology and Evolution*, 23(11), 2175–2190. <https://doi.org/10.1093/molbev/msl089>
- Downie, S. R., & Jansen, R. K. (2015). A comparative analysis of whole plastid genomes from the Apiales: Expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. *Systematic Botany*, 40(1), 336–351. <https://doi.org/10.1600/036364415X686620>
- Fry, A. J., & Wernegreen, J. J. (2005). The roles of positive and negative selection in the molecular evolution of insect endosymbionts. *Gene*, 355, 1–10. <https://doi.org/10.1016/j.gene.2005.05.021>
- Greiner, S., Lehwark, P., & Bock, R. (2019). Organellar Genome DRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, 47(W1), W59–W64. <https://doi.org/10.1093/nar/gkz238>
- Grewe, F., Guo, W., Gubbels, E. A., Hansen, A. K., & Mower, J. P. (2013). Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evolutionary Biology*, 13, 8. <https://doi.org/10.1186/1471-2148-13-8>
- Guisinger, M. M., Kuehl, J. V., Boore, J. L., & Jansen, R. K. (2011). Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: Rearrangements, repeats, and codon usage. *Molecular Biology and Evolution*, 28(1), 583–600. <https://doi.org/10.1093/molbev/msq229>
- Guo, W.-H., Grewe, F., Cobo-Clark, A., Fan, W.-S., Duan, Z.-L., Adams, R. P., Schwarzbach, A., & Mower, J. P. (2014). Predominant and sub-stoichiometric isomers of the plastid genome coexist within *Juniperus* plants and have shifted multiple times during Cupressophyte evolution. *Genome Biology and Evolution*, 6(3), 580–590. <https://doi.org/10.1093/gbe/evu046>
- Hall, B. G. (2005). Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Molecular Biology and Evolution*, 22(3), 792–802. <https://doi.org/10.1093/molbev/msi066>
- Heinhorst, S., & Cannon, G. C. (1993). DNA replication in chloroplasts. *Journal of Cell Science*, 104(1), 1–9. <https://doi.org/10.1083/jcb.120.2.557>
- Hirao, T., Watanabe, A., Kurita, M., Kondo, T., & Takata, K. (2008). Complete nucleotide sequence of the *Cryptomeria japonica* D. Don. chloroplast genome and comparative chloroplast genomics: diversified genomic structure of coniferous species. *BMC Plant Biology*, 8(1), 70. <https://doi.org/10.1186/1471-2229-8-70>
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17, 754–755. <https://doi.org/10.1093/bioinformatics/17.8.754>
- IBM Corporation (2014). *SPSS Statistics (version 22)*. IBM Corporation.
- Jin, D.-M., Wicke, S., Gan, L., Yang, J.-B., Jin, J.-J., & Yi, T.-S. (2020). The loss of the inverted repeat in the putranjivoid clade of Malpighiales. *Frontiers in Plant Science*, 11, 942. <https://doi.org/10.3389/fpls.2020.00942>
- Julian, T. F., Nevill, P. G., Dixon, K., & Small, I. (2017). What can we do with 1000 plastid genomes? *Plant Journal*, 90(4), 801–818. <https://doi.org/10.1111/tjp.13491>
- Karnkowska, A., Bennett, M. S., & Triemer, R. E. (2018). Dynamic evolution of inverted repeats in Euglenophyta plastid genomes. *Scientific Reports*, 8, 16071. <https://doi.org/10.1038/s41598-018-34457-w>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>
- Khakhlova, O., & Bock, R. (2006). Elimination of deleterious mutations in plastid genomes by gene conversion. *The Plant Journal*, 46(1), 85–94. <https://doi.org/10.1111/j.1365-313X.2006.02673.x>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35(6), 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Kwon, E. C., Kim, J. H., & Kim, N. S. (2020). Comprehensive genomic analyses with 115 plastomes from algae to seed plants: Structure, gene contents, GC contents, and introns. *Genes and Genomics*, 42(5), 553–570. <https://doi.org/10.1007/s13258-020-00923-x>



- Lanfear, R., Ho, S. Y. W., Davies, T. J., Moles, A. T., Aarssen, L., Swenson, N. G., Warman, L., Zanne, A. E., & Allen, A. P. (2013). Taller plants have lower rates of molecular evolution. *Nature Communications*, 4, 1879. <https://doi.org/10.1038/ncomms2836>
- Lanfear, R., Welch, J. J., & Bromham, L. (2010). Watching the clock: Studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution*, 25(9), 495–503. <https://doi.org/10.1016/j.tree.2010.06.007>
- Lemieux, C., Otis, C., & Turmel, M. (2014). Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics*, 15, 857. <https://doi.org/10.1186/1471-2164-15-857>
- Lemieux, C., Otis, C., & Turmel, M. (2016). Comparative chloroplast genome analyses of streptophyte green algae uncover major structural alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae. *Frontiers in Plant Science*, 7, 697. <https://doi.org/10.3389/fpls.2016.0069>
- Li, F.-W., Kuo, L. Y., Pryer, K. M., & Rothfels, C. J. (2016). Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Genome Biology and Evolution*, 8(8), 2452–2458. <https://doi.org/10.1093/gbe/evw167>
- Li, J., Gao, L., Chen, S., Tao, K., Su, Y.-J., & Wang, T. (2016). Evolution of short inverted repeat in Cupressophytes, transfer of *AccD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of *Sciadopityaceae*. *Scientific Reports*, 6, 20934. <https://doi.org/10.1038/srep20934>
- Li, Z., De La Torre, A. R., Sterck, L., Cánovas, F. M., Avila, C., Merino, I., Cabezas, J. A., Cervera, M. T., Ingvarsson, P. K., & Van de Peer, Y. (2017). Single-copy genes as molecular markers for phylogenomic studies in seed plants. *Genome Biology and Evolution*, 9(5), 1130–1147. <https://doi.org/10.1093/gbe/evx070>
- Lin, C. P., Huang, J. P., Wu, C. S., Hsu, C. Y., & Chaw, S. M. (2010). Comparative chloroplast genomics reveals the evolution of Pinaceae genera and subfamilies. *Genome Biology and Evolution*, 2, 504–517. <https://doi.org/10.1093/gbe/evq036>
- Lin, C. P., Wu, C. S., Huang, Y. Y., & Chaw, S. M. (2012). The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biology and Evolution*, 4(3), 374–381. <https://doi.org/10.1093/gbe/evs021>
- Maréchal, A., & Brisson, N. (2010). Recombination and the maintenance of plant organelle genome stability. *New Phytologist*, 186(2), 299–317. <https://doi.org/10.1111/j.1469-8137.2010.03195.x>
- Mathews, S. (2009). Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. *American Journal of Botany*, 96(1), 228–236. <https://doi.org/10.1207/27793085>
- McCoy, S. R., Kuehl, J. V., Boore, J. L., & Raubeson, L. A. (2008). The complete plastid genome sequence of *Welwitschia mirabilis*: An unusually compact plastome with accelerated divergence rates. *BMC Evolutionary Biology*, 8(1), 130. <https://doi.org/10.1186/1471-2148-8-130>
- Moore, M. J., Dhingra, A., Soltis, P. S., Shaw, R., Farmerie, W. G., Foltá, K. M., & Soltis, D. E. (2006). Rapid and accurate pyrosequencing of angiosperm plastid genomes. *BMC Plant Biology*, 6, 17. <https://doi.org/10.1186/1471-2229-6-17>
- Mower, J. P., & Vickrey, T. L. (2018). Chapter nine—structural diversity among plastid genomes of land plants. S. M. Chaw, & R. K. Jansen (eds), *Plastid genome evolution* (pp. 263–292). Academic Press.
- Pacheco, T. G., Lopes, A. S., Viana, G. D. M., Silva, O. N., Silva, G. M., Vieira, L. N., Guerra, M. P., Nodari, R. O., Souza, E. M., Pedrosa, F. O., Otoni, W. C., & Rogalski, M. (2019). Genetic, evolutionary and phylogenetic aspects of the plastome of annatto (*Bixa orellana* L.), the Amazonian commercial species of natural dyes. *Planta*, 249(2), 563–582. <https://doi.org/10.1007/s00425-018-3023-6>
- Palmer, J. D., Soltis, D. E., & Chase, M. W. (2004). The plant tree of life: An overview and some points of view. *American Journal of Botany*, 91(10), 1437–1445. <https://doi.org/10.3732/ajb.91.10.1437>
- Palmer, J. D., & Thompson, W. F. (1981). Rearrangements in the chloroplast genomes of mung bean and pea. *Proceedings of the National Academy of Sciences*, 78(9), 5533–5537. <https://doi.org/10.1073/pnas.78.9.5533>
- Palmer, J. D., & Thompson, W. F. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*, 29(2), 537–550. [https://doi.org/10.1016/0092-8674\(82\)90170-2](https://doi.org/10.1016/0092-8674(82)90170-2)
- Perry, A. S., & Wolfe, K. H. (2002). Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *Journal of Molecular Evolution*, 55(5), 501–508. <https://doi.org/10.1007/PL00020998>
- Ping, J.-Y., Feng, P.-P., Hao, J., Li, J.-Y., Su, Y.-J., & Wang, T. (2021). The molecular evolution pattern of *rps12* gene in gymnosperms (in Chinese). *Chin Science Bulletin*, 66(24), 3182–3193. <https://doi.org/10.1360/TB-2020-1528>
- Ping, J.-Y., Li, A.-M., Feng, P.-P., Zhu, M., Su, Y. J., & Wang, T. (2021). The highly conserved *rps12* gene in ferns provides strong evidence for decreased substitution rates in the inverted repeat region. *Plant Systematics and Evolution*, 307(2), 26. <https://doi.org/10.1007/s00606-021-01750-7>
- Pond, S. L., Frost, S. D., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, 21, 676–679. <https://doi.org/10.1093/bioinformatics/bti079>
- Ran, J.-H., Shen, T.-T., Wang, M.-M., & Wang, X.-Q. (2018). Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. *Proceedings of the Royal Society B-Biological Sciences*, 285, 20181012. <https://doi.org/10.1098/rspb.2018.1012>
- Ruhlman, T. A., & Jansen, R. K. (2014). The plastid genomes of flowering plants. In P. Maliga (Ed.), *Chloroplast biotechnology: Methods and protocols* (pp. 3–38). Humana Press.
- Ruhlman, T. A., Zhang, J., Blazier, J. C., Sabir, J. S. M., & Jansen, R. K. (2017). Recombination-dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *American Journal of Botany*, 104(4), 559–572. <https://doi.org/10.3732/ajb.1600453>
- Sanderson, M. J., Copetti, D., Burquez, A., Bustamante, E., Charboneau, J. L. M., Eguiarte, L. E., Kumar, S., Lee, H. O., Lee, J., McMahon, M., Steele, K., Wing, R., Yang, T.-J., Zwickl, D., & Wojciechowski, M. F. (2015). Exceptional reduction of the plastid genome of *saguaro cactus* (*Carnegiea gigantea*): Loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany*, 102(7), 1115–1127. <https://doi.org/10.3732/ajb.1500184>
- Shaw, J., Shafer, H. L., Leonard, O. R., Kovach, M. J., Schorr, M., & Morris, A. B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: The tortoise and the hare IV. *American Journal of Botany*, 101, 1987–2004. <https://doi.org/10.3732/ajb.1400398>
- Stamatakis, A. (2014). RaxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- Sun, Y.-X., Moore, M. J., Meng, A.-P., Soltis, P. S., Soltis, D. E., Li, J.-Q., & Wang, H.-C. (2013). Complete plastid genome sequencing of *Trochodendraceae* reveals a significant expansion of the inverted repeat and suggests a Paleogene divergence between the two extant species. *PLoS One*, 8(4), e60429. <https://doi.org/10.1371/journal.pone.0060429>
- Swofford, D. L. (2002). *PAUP\*: Phylogenetic analysis using parsimony (and other methods)*, version 4. Sinauer Associates.
- Turmel, M., Otis, C., & Lemieux, C. (2015). Dynamic evolution of the chloroplast genome in the green algal classes *Pedinophyceae* and

- Trebouxiophyceae. *Genome Biology and Evolution*, 7, 2062–2082. <https://doi.org/10.1093/gbe/evv130>
- Wang, B., Jiang, B., Zhou, Y., Su, Y.-J., & Wang, T. (2015). Higher substitution rates and lower  $dN/dS$  for the plastid genes in Gnetales than other gymnosperms. *Biochemical Systematics and Ecology*, 59, 278–287. <https://doi.org/10.1016/j.bse.2015.02.009>
- Wang, J.-H., Moore, M. J., Wang, H., Zhu, Z.-X., & Wang, H.-F. (2020). Plastome evolution and phylogenetic relationships among malvaceae subfamilies. *Gene*, 765, 145103. <https://doi.org/10.1016/j.gene.2020.145103>
- Wicke, S., Schaferhoff, B., de Pamphilis, C. W., & Muller, K. F. (2014). Disproportional plastome-wide increase of substitution rates and relaxed purifying selection in genes of carnivorous Lentibulariaceae. *Molecular Biology and Evolution*, 31(3), 529–545. <https://doi.org/10.1093/molbev/mst261>
- Wicke, S., Schneeweiss, G. M., de Pamphilis, C. W., Muller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, 76, 273–297. <https://doi.org/10.1007/s11103-011-9762-4>
- Wolfe, K. H., Li, W. H., & Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences*, 84, 9054–9058. <https://doi.org/10.1073/pnas.84.24.9054>
- Wu, C. S., & Chaw, S. M. (2014). Highly rearranged and size-variable chloroplast genomes in conifers II clade (cupressophytes): Evolution towards shorter intergenic spacers. *Plant Biotechnology Journal*, 12(3), 344–353. <https://doi.org/10.1111/pbi.12141>
- Wu, C. S., & Chaw, S. M. (2015). Evolutionary stasis in Cycad plastomes and the first case of plastome GC-biased gene conversion. *Genome Biology and Evolution*, 7, 2000–2009. <https://doi.org/10.1093/gbe/evv125>
- Wu, C. S., Lai, Y. T., Lin, C. P., Wang, Y. N., & Chaw, S. M. (2009). Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: Selection toward a lower-cost strategy. *Molecular Phylogenetics and Evolution*, 52(1), 115–124. <https://doi.org/10.1016/j.ympev.2008.12.026>
- Wu, C. S., Lin, C. P., Hsu, C. Y., Wang, R. J., & Chaw, S. M. (2011). Comparative chloroplast genomes of Pinaceae: Insights into the mechanism of diversified genomic organizations. *Genome Biology and Evolution*, 3, 309–319. <https://doi.org/10.1093/gbe/evr026>
- Wu, C. S., Wang, Y. N., Hsu, C. Y., Lin, C. P., & Chaw, S. M. (2011). Loss of different repeat copies from chloroplast genome of Pinaceae and Cupressophytes and influence of heterotachy on the evaluation of Gymnosperm phylogeny. *Genome Biology and Evolution*, 3(6), 1284–1295. <https://doi.org/10.1093/gbe/evr095>
- Wyman, S. K., Jansen, R. K., & Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, 20(17), 3252–3255. <https://doi.org/10.1093/bioinformatics/bth352>
- Yang, Z. (2004). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8), 1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Yang, Z. (2006). *Computational molecular evolution*. Oxford University Press.
- Yi, X., Gao, L., Wang, B., Su, Y.-J., & 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 Biology and Evolution*, 5(4), 688–698. <https://doi.org/10.1093/gbe/evt042>
- Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for *de novo* short read assembly using *de Bruijn* graphs. *Genome Research*, 18(5), 821–829. <https://doi.org/10.1101/gr.074492.107>
- Zhang, Y.-Z., Ma, J., Yang, B.-X., Li, R.-Y., Zhu, W., Sun, L.-L., Tian, J.-K., & Zhang, L. (2014). The complete chloroplast genome sequence of *Taxus chinense* var *mairei* (Taxaceae): Loss of an invert repeat region and comparative analysis with related species. *Gene*, 540(2), 201–209. <https://doi.org/10.1016/j.gene.2014.02.037>
- Zheng, W.-J., & Fu, L.-G. (1978). Cupressaceae. In Z.-Y. Wu (Ed.), *Flora of China*, Vol. 7. Science Press (in Chinese).
- Zhu, A.-D., Guo, W.-H., Gupta, S., Fan, W.-S., & Mower, J. P. (2016). Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytologist*, 209(4), 1747–1756. <https://doi.org/10.1111/nph.13743>

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Ping, J., Hao, J., Li, J., Yang, Y., Su, Y., & Wang, T. (2022). Loss of the IR region in conifer plastomes: Changes in the selection pressure and substitution rate of protein-coding genes. *Ecology and Evolution*, 12, e8499. <https://doi.org/10.1002/ece3.8499>