Ecology and Evolution

WILEY

RESEARCH ARTICLE OPEN ACCESS

Sequencing of Complete Chloroplast Genomes and Comparative Analyses of Codon Usage Bias and Phylogeny in *Globba* Species

Wenhui Liao | Xianlin Rong | Chunhai Wei | Mingli Zhang | Jianshuang Shen

Hangzhou Polytechnic University, Hangzhou, China

Correspondence: Jianshuang Shen (shenjianshuang18@163.com)

Received: 4 January 2026 | **Revised:** 15 March 2026 | **Accepted:** 19 March 2026

Keywords: chloroplast genome | codon usage bias | *Globba* | phylogeny

ABSTRACT

The genus *Globba* (Zingiberaceae) comprises over 100 species distributed worldwide, with a primary concentration in Southeast Asia. These plants are of significant interest due to their distinctive and delicate floral morphology as well as their potential medicinal value. In this study, we assembled and characterized the complete chloroplast (cp) genomes of three species: *Globba racemosa, G. "Mermaid Princess,"* and *G. "Ziyan"*. Each genome displayed the typical angiosperm quadripartite structure, with total lengths of 163,141 bp, 163,487 bp, and 163,535 bp, respectively. Genome annotation identified 133–134 unique genes, comprising 87 protein-coding genes, 38–39 tRNAs, and 8 rRNAs. Annotation and comparative analysis of a total of 11 *Globba* chloroplast genomes revealed that the *trnG-GCC* gene was present in two copies exclusively in *G. multiflora* and *G. racemosa*, while the *trnQ-UUG* gene was specifically absent from the chloroplast genome of *G. atrosanguinea*. Comparative genomic analysis indicated high synteny across all genomes, with no large-scale structural rearrangements detected; however, variations at the inverted repeat (IR) boundaries involving the *ycf1* and *ndhF* genes were observed. Codon usage bias analysis demonstrated a strong preference for A/U-ending codons (RSCU > 1.0), with leucine (Leu), threonine (Thr), and arginine (Arg) being the most frequently encoded amino acids, and 17 optimal codons were identified. Simple sequence repeat (SSR) analysis showed that the total number of SSRs per genome ranged from 60 to 70. Combined analyses of ENC-plot, PR2-plot, Neutral-plot, and BUSTED indicated that codon usage patterns in *Globba* chloroplast genomes are primarily shaped by mutation pressure. Phylogenetic reconstruction based on 11 complete chloroplast genome sequences revealed that all species were monophyletically grouped into two primary clades. These newly generated cp genomes provide valuable genomic resources for enhancing phylogenetic resolution, elucidating molecular evolutionary patterns, informing conservation strategies, and supporting further research on ornamental and medicinal plants within *Globba*.

1 | Introduction

The genus *Globba* L. (Zingiberaceae) is a species-rich and ecologically significant lineage within the subfamily Zingiberoideae, tribe Globbeae, primarily distributed across tropical Asia (Kress et al. 2002; Williams et al. 2004). Beyond its systematic importance, *Globba* species are valued for their ornamental appeal and medicinal potential (Li et al. 2018). However, the taxonomy and evolutionary history of the genus remain poorly resolved.

Phylogenetic studies utilizing ITS and chloroplast *trnK–matK* sequences have revealed that traditional infrageneric classifications—based primarily on stamen appendage number—are phenetically informative but do not reflect monophyletic

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.

© 2026 The Author(s). *Ecology and Evolution* published by British Ecological Society and John Wiley & Sons Ltd.

groups, suggesting multiple independent origins and repeated evolutionary transitions of this character (Kuncharoen et al. 2018). Molecular evidence further indicates that genera such as *Mantisia* are nested within *Globba*, necessitating a revised classification that integrates morphological traits (e.g., inflorescence architecture, fruit morphology, stamen appendage structure) with molecular data (Sangvirotjanapat et al. 2019).

At the species level, high morphological similarity among many *Globba* species complicates specimen identification and blurs species boundaries. Recent taxonomic revisions underscore the critical role of detailed comparisons of living and preserved specimens, coupled with micromorphological analysis, for accurate species delimitation. Reproductive traits, such as andromonoecy, may also provide key insights into pollination evolution and species circumscription, highlighting the need to integrate fine-scale morphology and reproductive biology into systematic studies (Sangvirotjanapat et al. 2019).

Chloroplast (cp) genomes have emerged as powerful tools for species discrimination and phylogenetic reconstruction in Zingiberoideae due to their structural conservation and ease of comparison (Li et al. 2021). Comparative cp genomic analyses across several genera, including *Globba*, have identified hypervariable regions suitable for developing specific barcodes or molecular markers, which are particularly useful for distinguishing morphologically cryptic lineages and resolving relationships at the population or closely related species level (Li et al. 2021). Despite these advances, a comprehensive analysis of cp genome evolution in *Globba*—encompassing structural variation, codon usage bias, and patterns of natural selection across multiple species—is still lacking.

In addition to its systematic interest, *Globba* possesses notable chemical and medicinal properties (Menon and Dan 2009; Naznin et al. 2019; Li et al. 2017). Studies on *Globba schomburgkii* have analyzed the volatile chemical composition of extracts from different plant parts, revealing antibacterial and antifungal activities (Doungchawee et al. 2019). Despite the traditional medicinal uses of *Globba* species (e.g., *G. schomburgkii*) for treating conditions such as gastric disorders and inflammation, and preliminary evidence of their antibacterial and antifungal activities, systematic investigations into their bioactive constituents and corresponding pharmacological mechanisms remain notably limited (Li et al. 2017). Furthermore, *Globba* plants host diverse endophytic microorganisms with bioprospecting potential (Matsumoto and Takahashi 2019). Notable examples are the endophytic actinomycete *Micromonospora globbae*, which was isolated from the roots of *Globba winitii* (Kuncharoen et al. 2018; Santra and Bera 2025).

G. "Mermaid Princess" and *G.* "Ziyan" are cultivated *Globba* varieties introduced from Thailand. They hold significant economic value, primarily as cut flowers, and are currently being promoted in regions such as Guangzhou and Shanghai, China. However, studies on their genetic background and taxonomic classification remain limited. Although the complete chloroplast genome of *G. racemosa* from Quzhou, China, has been previously reported by our team, the taxonomic status of *G. racemosa* in Quzhou has been controversial (Shen et al. 2026). Several reports suggest that the Zhejiang population may represent a distinct cryptic species or a new taxon (Qian et al. 2013). To clarify its phylogenetic position and resolve its relationship with the *G. racemosa* (Enshi population), we re-sequenced the chloroplast genome of *G. racemosa* from Enshi, China.

In this study, we present newly sequenced chloroplast genomes of "Mermaid Princess", "Ziyan", and the Quzhou population of *G. racemosa*. These data aim to fill existing knowledge gaps and enhance our understanding of germplasm resources within the genus *Globba*. To address the persistent taxonomic uncertainties and evolutionary questions surrounding *Globba*, this study presents a comprehensive comparative analysis of complete chloroplast genomes from multiple representative species. By examining genome structure, codon usage patterns, and signatures of natural selection, we aim to (1) establish a robust phylogenetic framework for the genus, (2) elucidate the evolutionary mechanisms underlying cp genome diversity, and (3) develop reliable molecular markers for species identification. The findings are expected to contribute significantly to the systematics of *Globba* and to support future efforts in conservation and bioprospecting within this economically and ecologically important genus.

2 | Materials and Methods

2.1 | Plant Material Sampling

Fresh and healthy leaves of four *Globba* species (including *Globba racemosa* Smith collected from Enshi, China, *G.* "Mermaid Princess" collected from Thailand, and *G.* "Ziyan" collected from Thailand) were planted from the Hangzhou Polytechnic University, located at 120.391°E, 30.299°N in Hangzhou, China (Figure 1). Voucher specimens of *G. racemosa* (collected from Enshi, China), G. "Mermaid Princess" (Thailand), and G. "Ziyan" (Thailand) were deposited at the Institute of Botany, Chinese Academy of Sciences, and the Herbarium of Sun Yat-sen University under the accession numbers 12637771, 18332029, and 18332019, respectively. Specimen data are accessible via the Chinese Virtual Herbarium (https://www.cvh.ac.cn/index.php, cvh@ibcas.ac.cn) and Plant Science Data Center (https://www.iplant.cn/, bigdata@ibcas.ac.cn). The collected leaf samples were immediately frozen in liquid nitrogen and stored at −80°C.

2.2 | DNA Extraction, Chloroplast Genome Sequencing and Assembly

The total genomic DNA was extracted from these leaves using the improved CTAB method (Doyle 1987). Genomic DNA samples were subjected to ultrasonic fragmentation to generate fragments of appropriate size. The resulting fragments underwent end repair, followed by the addition of an "A" base at the 3′ ends and subsequent ligation of sequencing adapters. A preliminary test was conducted to assess the ligation efficiency. Thereafter, fragment selection and PCR amplification were carried out to enrich for properly constructed library molecules. The constructed libraries were then purified to remove adapter contaminants



FIGURE 1 | The reference image of the plant of *Globba* species. (A) *Globba racemosa* Smith (collected from Enshi, China); (B) *Globba* "Mermaid Princess"; (C) *Globba* "Ziyan".

and other impurities. Finaly, library quality was verified prior to high-throughput sequencing. The qualified libraries were subjected to sequencing on the BGISEQ-T7 platform.

All raw sequencing reads were first subjected to quality filtering and adapter trimming using Fastp (Chen et al. 2018), to remove low-quality bases and residual adapter sequences. The resulting clean reads were then aligned to a custom-built chloroplast genome database (constructed by Bio&Data Biotechnologies Inc., Guangzhou, China) with Minimap2 v2.30 (Li 2021). De novo assembly of the chloroplast genome was carried out using NOVOPlasty, with the published chloroplast genome of *Zingiber officinale* (GenBank accession number: MH161428) designated as the reference sequence. The initial assembly was subsequently verified and optimized using Geneious v.9.1.8 software (Kearse et al. 2012), followed by manual adjustments to achieve a complete, circularized genome sequence.

The assembled chloroplast genome was annotated using the web-based tool GeSeq for identifying protein-coding genes and ribosomal RNAs, in conjunction with tRNAscan-SE for the prediction of transfer RNA genes (Tillich et al. 2017; Lowe and Edd 1997). All automated annotations were subsequently manually curated using Geneious to correct gene boundaries and intron/exon positions, yielding the final annotated genome. A circular gene map was subsequently generated using OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html). Drawing the structures of cis- and trans-splicing genes using CPGView (Liu et al. 2023).

2.3 | Comparative Genome Analysis

For comparative genomic analysis, two complementary visualization strategies were employed, both with *Globba racemosa*

(GenBank accession number: PX549197) as the reference sequence. First, multiple sequence alignment was conducted using the MAFFT v7.487 software (Katoh and Standley 2013); the alignment results were then subjected to linear visualization. In the linear plot, the height of each peak corresponds to the sequence identity at the respective site, while completely disconnected positions represent sequence gaps. For the second strategy, sequence alignment was performed using BLASTn v2.16.0+, followed by circular visualization of the alignment outcomes (Johnson et al. 2008). In this circular plot, the color depth indicates the level of sequence identity at each site, and blank regions correspond to sequence gaps. Furthermore, the expansion and contraction of the inverted repeat (IR) regions were illustrated using CPJSdraw (0.0.1) (Li et al. 2023). The two sequences (*G. variabilis* PP542022 and *G. atrosanguinea* PP542021) were excluded from the comparative analysis of inverted repeats, as they exhibited no long IR sequences but instead contained numerous short inverted repeats, a pattern inconsistent with the typical features of chloroplast genomes.

2.4 | Simple Sequence Repeat Analysis

Repetitive sequences in the chloroplast genome were comprehensively analyzed. Chloroplast simple sequence repeats (cpSSRs) were identified using MISA (MIcroSAtellite identification tool) software (Thiel et al. 2003; Beier et al. 2017) with the following thresholds: mononucleotide ≥ 10 repeats, dinucleotide ≥ 6 repeats, and tri- to hexanucleotide ≥ 5 repeats; the maximum interruption for compound SSRs was set to 100 bp.

2.5 | Nucleotide Diversity and Sliding Window Analysis

To assess the level of sequence polymorphism, the nucleotide diversity (Pi) was estimated for the multiple sequence alignment of the chloroplast genomes using DnaSP 6.0. A sliding window analysis was performed with a window size of 500 bp and a step size of 50 bp. Peaks of Pi values along the genome were subsequently mapped to specific genomic regions and gene loci to identify hypervariable hotspots.

2.6 | Codon Usage Bias Analysis

The relative synonymous codon usage (RSCU) for the coding genes of each species was using the RSCU-Plot web server (available at https://pcg-lab.shinyapps.io/RSCU-Plot/). Considering the presence of limited RNA-editing in start codons and a very small number of non-ATG start codons in chloroplast genomes, all non-ATG initiation codons were directly removed during the analysis. Correlation plots were generated using the R language software package (Sharp and Li 1986). The overall GC content (GC_{all}) and the GC content at the first, second, and third codon positions (GC_1, GC_2, GC_3) were computed using the EMBOSS online toolkit (http://www.bioinformatics.nl/emboss-explorer/). To evaluate effective number of codons (ENC) in our dataset, we used the chips tool from the EMBOSS suite.

ENC is a widely used metric to assess the overall level of codon usage bias. Using ENC as the ordinate (y-axis) and GC_3 as the abscissa (x-axis), an ENC-plot (ENC vs. GC_3) with the standard expected curve scatter was constructed. A standard curve was fitted based on the distribution of the scatter points (Wright 1990; Comeron and Aguadé 1998). The frequency of the ENC ratio was then statistically analyzed by calculating the distance between each scatter point and the standard curve. The codon usage bias was assessed according to the degree of deviation of the ENC values from the standard curve.

Neutrality analysis (GC_{12} vs. GC_3) involves plotting GC_{12} against GC_3 to explore the relationship between GC_{12} and GC_3 and evaluate the equilibrium between mutation and selection in shaping codon usage bias (CUB) (Sueoka 1988). In this plot, GC_3 is on the X-axis and GC_{12} is on the Y-axis, with each gene represented by a distinct point. Parity rule 2 (PR2) plot analysis was performed to investigate the combined effects of mutation and selection on codon usage bias (CUB). The plot was generated with $G_3/(G_3+C_3)$ values on the X-axis and $A_3/(A_3+U_3)$ values on the Y-axis, dividing the data into four quadrants in a scatter diagram (Sueoka 1988).

2.7 | Optimal Codon Analysis

The effective number of codons (ENC) was used as a reference standard for codon usage bias. All coding sequences (CDS) were ranked in descending order based on their ENC values. The top 10% and bottom 10% of genes were selected to construct high-bias and low-bias gene sets, respectively. The ΔRSCU value ($\Delta RSCU = RSCU_{high\text{-}bias} - RSCU_{low\text{-}bias}$) was then calculated for each codon. Codons with ΔRSCU ≥ 0.08 were defined as preferentially expressed codons. Additionally, codons with an RSCU value > 1 were classified as high-frequency codons. Codons satisfying both criteria ΔRSCU ≥ 0.08 and RSCU > 1 were identified as optimal codons.

2.8 | Phylogenetic Analysis

To elucidate the evolutionary relationships of the 11 samples within the genus *Globba*, a phylogenetic tree was reconstructed based on chloroplast genome sequences and eighteen genes with Pi values exceeding 0.65 (including *trnQ-UUG, rps16, atpH, atpI, rpoB, trnC-GCA, trnT-UGU, trnL-UAA, ndhC, trnV-UAC, accD, psal, psbE, petL, ycf1, ndhF, psaC*, and *ndhE*.), respectively. *Zingiber officinale* (MH161428) and *Zingiber zerumbet* (MK262726) as outgroup. The ML tree was constructed using IQ-TREE software under the TVM + F + R3 model, with support values estimated from 1000 bootstrap replicates.

2.9 | Detection of Positive Selection

The selection pressure on protein-coding genes was assessed using the BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification) model implemented in the HyPhy software suite. This model tests for gene-wide evidence of episodic positive selection (Kosakovsky Pond et al. 2020).

TABLE 1 | Characteristics of cp genomes of the *Globba* species.

Genome feature	*G. racemosa* Smith (Quzhou, China)	*G. racemosa* Smith (Enshi, China)	*G.* "Mermaid Princess"	*G.* "Ziyan"	*G. lancangensis*	*G. marantina*	*G. multiflora*	*G. schomburgkii*	*G. schomburgkii* var. angustata	*G. variabilis*	*G. atrosanguinea*
GenBank accession	PX430557	PX549197	PX549198	PX549199	MT473704*	MT473705*	MT473706*	MK262735*	MT473707*	PP542022*	PP542021*
Total length (bp)	163,321	163,141	163,487	163,535	163,306	162,774	163,199	163,325	163,432	165,482	161,896
Length of LSC (bp)	88,111	87,978	88,677	88,700	88,557	87,989	88,019	88,498	88,603	—	—
Length of SSC (bp)	15,770	15,719	15,498	15,529	15,393	15,487	15,718	15,525	15,523	—	—
Length of IR (bp)	29,720	29,722	29,656	29,653	29,678	29,649	29,731	29,651	29,653	—	—
Number of genes	134	134	133	133	133	133	134	133	133	133	132
Number of protein coding genes	87	87	87	87	87	87	87	87	87	87	87
Number of tRNA genes	39	39	38	38	38	38	39	38	38	38	37
Number of rRNA genes	8	8	8	8	8	8	8	8	8	8	8
Total GC content (%)	35.8	35.84	35.86	35.81	35.73	35.92	35.85	35.85	35.83	36.21	36.07
LSC GC content (%)	33.53	33.59	33.47	33.44	33.34	33.58	33.59	33.50	33.46	—	—
SSC GC content (%)	28.69	28.74	29.38	29.15	29.07	29.31	28.82	29.17	29.17	—	—
IR GC content (%)	41.03	41.05	41.11	41.11	41.03	41.11	41.04	41.12	41.12	—	—

*Sequence downloaded from Gen Bank, then gene annotation of the chloroplast genomes were obtained by annotating the chloroplast genome using GeSeq and tRNAscan-SE, followed by manual correction. The cp genomes of *G. variabilis* and *G. atrosanguinea* without completely complementary IR region.

TABLE 2 | Genes present in the 11 sequenced chloroplast genomes in *Globba*.

Photosynthesis	Subunits of photosystem I	*psaA, psaB, psaC, psaI, psaJ, pafI***, *pafII*
	Subunits of photosystem II	*psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ, pbf1*
	Subunits of cytochrome b/f complex	*petA, petB**, *petD**, *petG, petL, petN*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF**, *atpH, atpI*
	Subunits of NADH dehydrogenase	*ndhA**, *ndhB* (×2) *, *ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK*
	Subunit of rubisco	*rbcL*
Self-replication	RNA polymerase	*rpoA, rpoB, rpoC1**, *rpoC2*
	Large subunit of ribosomal proteins	*rpl2 (×2)**, *rpl14, rpl16**, *rpl20, rpl22, rpl23* (×2), *rpl32, rpl33, rpl36*
	Small subunit of ribosomal proteins	*rps2, rps3, rps4, rps7* (×2), *rps8, rps11, rps12* (×2)**, *rps14, rps15, rps16**, *rps18, rps19* (×2)
	Ribosomal RNAs	*rrn4.5* (×2), *rrn5* (×2), *rrn16* (×2), *rrn23* (×2)
	Transfer RNAs	*trnA-UGC* (×2) *, *trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnfM-CAU*, ①*trnG-GCC*, *trnG-UCC**, *trnH-GUG* (×2), *trnI-CAU* (×2), *trnI-GAU* (×2)*, *trnK-UUU**, *trnL-CAA* (×2), *trnL-UAA**, *trnL-UAG, trnM-CAU, trnN-GUU* (×2), *trnP-UGG*, ②*trnQ-UUG, trnR-ACG* (×2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA, trnS-UGA, trnT-GGU*, *trnT-UGU*, *trnV-GAC* (×2), *trnV-UAC**, *trnW-CCA*, *trnY-GUA*
Other genes	Subunit of acetyl-coA-carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Envelop membrane protein	*cemA*
	Protease	*clpP1***
	Translational initiation factor	*infA*
	Maturase	*matK*
Unknown function	Conserved open reading frames	*ycf1* (×2), *ycf2* (×2)

Note: (×2): gene with two copies; *: gene containing one intron; **: gene containing two introns; ①:*trnG-GCC* exist two gene copies only in the chloroplast genome of *G. multiflora* and *G. racemosa*; ② The *trnQ-UUG* gene was absent from the chloroplast genome of *G. atrosanguinea*.

TABLE 3 | Number and comparison of SSR distribution by gen regions of the chloroplast genome sequences across 11 *Globba* species.

		Comparison of SSR distribution by gene regions					
Sample	**Total_SSRs**	**LSC**	**SSC**	**IR**	**IGS**	**Intron**	**CDS**
G. schomburgkii (MK262735)	64	46	10	8	45	10	9
G. lancangensis (MT473704)	67	47	12	8	50	8	9
G. amarantina (MT473705)	68	53	9	6	51	9	8
G. multiflora (MT473706)	70	46	14	10	51	11	8
G. schomburgkii var. angustata (MT473707)	65	47	10	8	47	10	8
G. atrosanguinea (PP542021)	60	/	/	/	/	10	10
G. variabilis (PP542022)	62	/	/	/	/	8	10
G. racemosa (PX549197)	70	48	14	8	49	11	10
G. racemosa (PX430557)	68	47	13	8	51	8	9
G. "Mermaid Princess" (PX549198)	69	48	13	8	49	11	9
G. "Ziyan" (PX549199)	67	47	12	8	51	7	9

TABLE 4 | optimal codons in the chloroplast genomes of 11 *Globba* species.

Rank	Codon	AA	RSCU	ΔRSCU
1	UUA	Leu	1.933	1.433
2	ACU	Thr	1.547	1.287
3	CGU	Arg	1.342	1.285
4	TCU	Ser	1.739	0.759
5	GGU	Gly	1.393	0.607
6	AUU	Ile	1.477	0.561
7	CGU	Arg	1.266	0.506
8	GUA	Val	1.545	0.504
9	AAA	Lys	1.483	0.474
10	CCU	Pro	1.666	0.464
11	CAA	Gln	1.544	0.444
12	UUU	Phe	1.311	0.412
13	AAU	Asn	1.538	0.302
14	CAU	H	1.587	0.28
15	AGU	Ser	1.247	0.273
16	GAA	Glu	1.512	0.264
17	GCU	Ala	1.82	0.146

3 | Results

3.1 | Chloroplast Genome Characterization

The complete chloroplast genomes of three *Globba* species (including *Globba racemosa*, *G.* "Mermaid Princess", and *G.* "Ziyan") were sequenced and assembled. The chloroplast genomes are 163,141 bp (*G. racemosa* from Enshi, China), 163,487 bp ("Mermaid Princess" from Thailand), and 163,535 bp ("Ziyan" from Thailand), all exhibiting the conserved quadripartite structure (Figure 2, Table 1). The 12 genes, including *rps16, atpF, rpoC1, pafI, clpP1, petB, petD, rpl16, rpl2, ndhB* (*two* copies), *ndhA*, and *rpl2*, were cis-splicing genes in the cp genomes of three *Globba* species (Figure S1).

To elucidate the evolutionary dynamics of chloroplast genomes in *Globba*, eleven species were compared, including three newly sequenced in this study. A moderate variation in genome sizes was found, ranging from 161,896 bp (*G. atrosanguinea*) to 165,482 bp (*G. variabilis*). The chloroplast genomes of nine *Globba* species exhibit the typical quadripartite structure, comprising a large single-copy (LSC) region, a small single-copy (SSC) region, and a pair of inverted repeats (IRs). However, this structure was not fully conserved in *G. variabilis* and *G. atrosanguinea*, as they exhibited no long IR sequences but instead contained numerous short inverted repeats.

Notably, the number of protein-coding genes was invariant at 87, the rRNA gene count remained constant at 8, while the total gene count varied from 132 to 134 due to differences in tRNA gene number (37–39). This variation was attributed to the lack of either the *trnG-GCC* or *trnQ-UUG* gene. Therefore, the genomic differences among these accessions are confined to variations in tRNA gene complement. Although the overall GC content was conserved (35.73%–36.21%), lineage-specific shifts were observed in the regional GC content of the LSC, SSC, and IR regions (Table 1).

The genes were categorized according to their putative functions into three groups: those involved in photosynthesis, self-replication, and other processes, which are summarized in Table 2. Among these, 15 genes contain a single intron, while three genes (*psfI*, *rps12* and *clpP1*) possess double introns (Table 2). The *trnG-GCC* gene was present in two copies exclusively in the chloroplast genomes of *G. multiflora* and *G. racemosa*, while the *trnQ-UUG* gene was absent specifically from the chloroplast genome of *G. atrosanguinea*.

3.2 | Comparative and Nucleotide Diversity Analysis of Chloroplast Genomes

The chloroplast genomes of *Globba* species exhibit high structural conservation and collinearity (Figure 3). Comparative visualization analyses demonstrated a conserved structure, with no evidence of gene inversions or genomic rearrangements. As shown in Figure 4, the chloroplast genomes of nine *Globba* species exhibit highly conserved boundary architecture, with no significant expansions or contractions. In *G. lancangensis*, the *ndhF* gene uniquely spanned the IRb/SSC junction, extending 17 bp into the IRb region. Notably, the two copies of the *ycf1* gene extend asymmetrically across the IR/SSC junctions, with expansions of 1555–1583 bp from IRa and 1–31 bp from IRb into the SSC, respectively. This phenomenon was observed in nine species but except in *G. variabilis* and *G. atrosanguinea* (Figure 4).

The analysis of nucleotide diversity (pi) across the *Globba* chloroplast genomes revealed substantial heterogeneity in sequence variation, characterized by conserved backgrounds interspersed with localized hypervariable regions (Figure 5). The genome-wide pi values range from 0 to 1, indicating the existence of sequence variation. Eighteen genes with Pi values exceeding 0.65 were identified. These genes are located within the LSC and SSC regions, including *trnQ-UUG*, *rps16*, *atpH*, *atpI*, *rpoB*, *trnC-GCA*, *trnT-UGU*, *trnL-UAA*, *ndhC*, *trnV-UAC*, *accD*, *psal*, *psbE*, *petL*, *ycf1*, *ndhF*, *psaC*, and *ndhE*.

3.3 | Distribution of Simple Sequence Repeats in *Globba* Chloroplast Genomes

Simple sequence repeats (SSRs) were systematically characterized across eleven *Globba* chloroplast genomes to elucidate their distribution patterns (Table 3). The total number of SSRs ranged from 60 (*G. atrosanguinea*) to 70 (*G. multiflora* and one *G. racemosa* from Enshi, China), reflecting moderate interspecific variation in SSR abundance. A conserved distribution pattern was observed across genomic regions: the large single-copy (LSC) region exhibited the highest SSR density (46–53), substantially exceeding that in the small single-copy (SSC, 9–14) and inverted repeat (IR, 6–10) regions. When classified by genomic context,

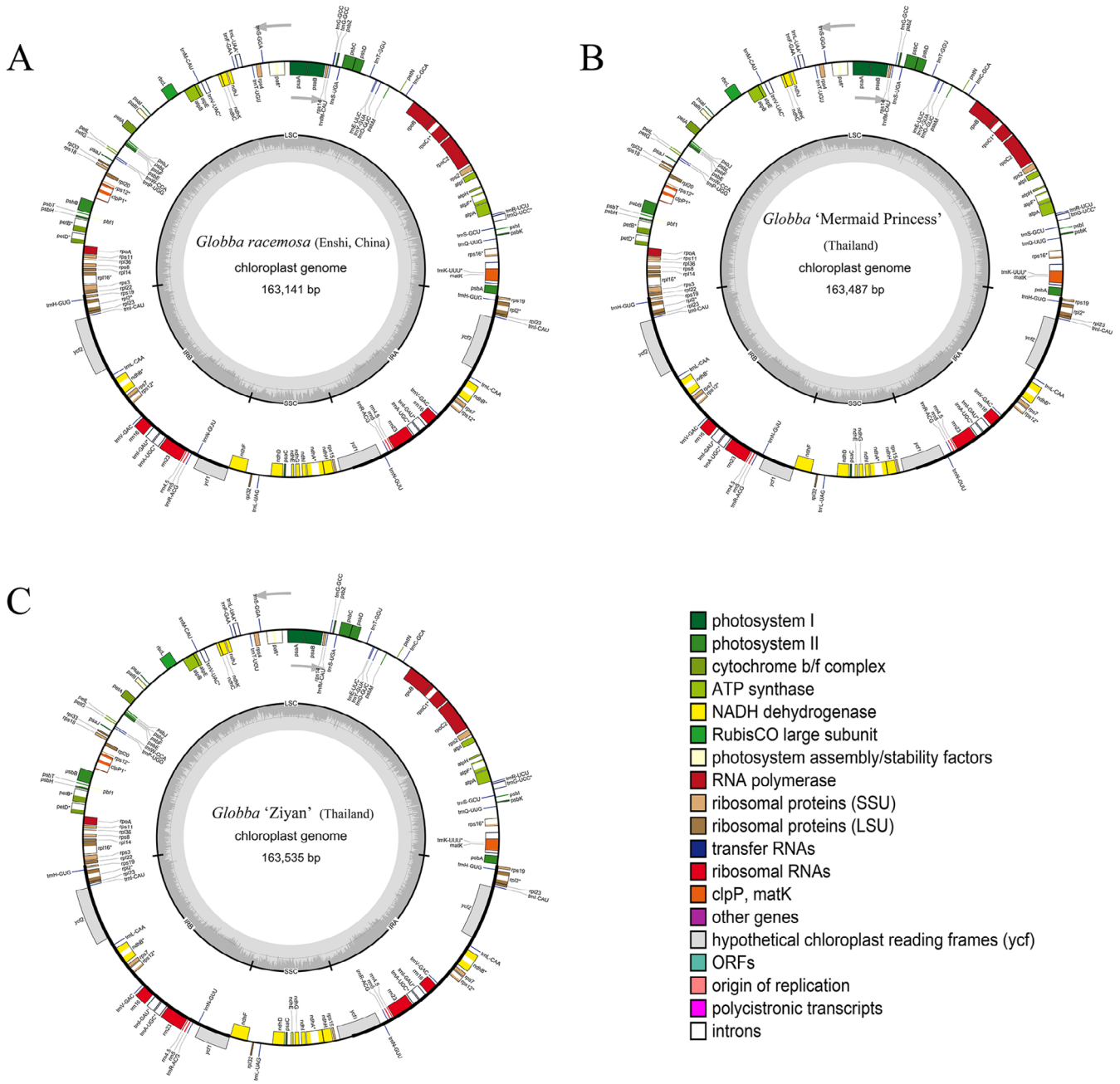


FIGURE 2 | Chloroplast genome maps of *Globba*. *G. racemosa* (Enshi, Zhejiang, China) (A), *G.* "Mermaid Princess" (Thailand) (B), and *G.* "Ziyan" (Thailand) (C). Genes on the outside of the map are transcribed in the clockwise direction and genes on the inside of the map are transcribed in the counterclockwise direction. The dark and light gray bars in the inner circle denote G + C and A + T contents, respectively.

SSRs were predominantly located in intergenic spacers (IGS, 45–51), with fewer occurrences in introns (7–11) and coding sequences (CDS, 8–10). These findings collectively demonstrate significant heterogeneity in SSR distribution across *Globba* chloroplast genomes, marked by a pronounced enrichment in non-coding regions and species-specific variation in both abundance and genomic localization.

3.4 | Codon Usage Bias of Chloroplast Genomes of *Globba* Species

A total of 87 CDS from *G. lancangensis* and *G. marantina*, 86 CDS from the other 9 species of *Globba*, were analyzed (Table S1, Figure 6, Figure S2). The mean contents of GC_1, GC_2, and GC_3 were 45.23%, 39.07%, and 27.26%, respectively (Table S1). The analysis identified 30 codons with a usage bias (RSCU > 1) in the chloroplast genomes of 11 *Globba* species, with the majority (29) terminating in either U (16 codons) or A (13 codons). This result demonstrates a pronounced overall preference for U- and A-ended synonymous codons in these genomes (Table S2).

Codon usage bias was analyzed in the eleven *Globba* chloroplast genomes, through which 17 optimal codons exhibiting both high relative synonymous codon usage (RSCU > 1) and high ΔRSCU values (ΔRSCU ≥ 0.08) were identified (Table 4). The strongest preference was shown by the codon UUA (Leu), with an RSCU of 1.933 and ΔRSCU of 1.433, followed by ACU (Thr,

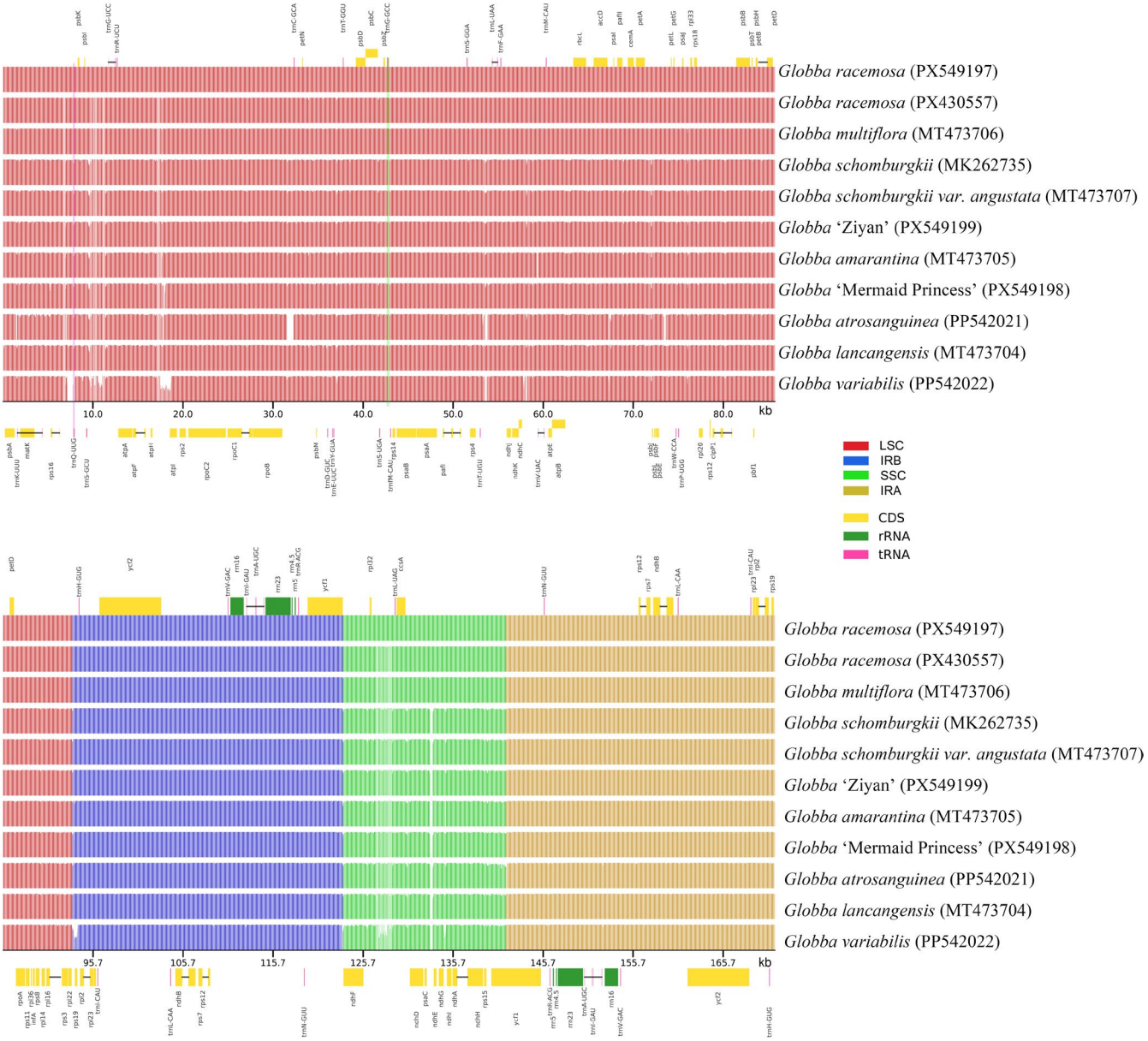


FIGURE 3 | Comparative visualization of chloroplast genome sequences across 11 *Globba* species. The y-axis represents sequence identity ranging from 0% to 100%, and the x-axis shows the position within the chloroplast genome. Using *G. racemosa* (Enshi, China) PX549197 as the reference sequence. The locations of *trnQ*-UUG and *trnG*-GCC are marked by purple and green vertical lines, respectively. Gaps are indicated by dots.

RSCU = 1.547, ΔRSCU = 1.287) and CGU (Arg, RSCU = 1.342, ΔRSCU = 1.285). These optimal codons were found to be distributed across multiple amino acids, including leucine, threonine, arginine, serine, and glycine, with a notable predominance of A/U-ending codons being observed.

3.5 | ENC-Plot, PR2-Plot, Neutral-Plot and BUSTED Analysis of Chloroplast Genomes of *Globba* Species

All protein-coding genes in the chloroplast genomes of the 11 studied *Globba* species exhibited Effective Number of Codons (ENC) values ranging from 26.75 to 60.58, with a mean of 46.00 (Table S1). The genes were categorized into two groups: a High Expression Group, comprising 95 genes (ENC: 26.75–38.86), and a Low Expression Group, comprising another 95 genes (ENC: 52.78–60.58) (Figure 7). Among these, only the *psaC* gene displayed an ENC value greater than 60, whereas the *rpl32* and *psbJ* genes had ENC values below 30. The proportion of genes plotting above the expected ENC-plot curve varied among species. *G. schomburgkii*, *G.* "*Mermaid Princess*", and *G.* "*Ziyan*" showed higher proportions, each exceeding 19%. In contrast, *G. racemosa*, *G. multiflora*, and *G. variabilis* had lower proportions, all below 17%. Notably, a clear difference was observed between the *G. racemosa* accessions from Quzhou and Enshi. In the Enshi accession, 16.1% of the genes were distributed above the expected curve, compared to 14.9% in the Quzhou accession.

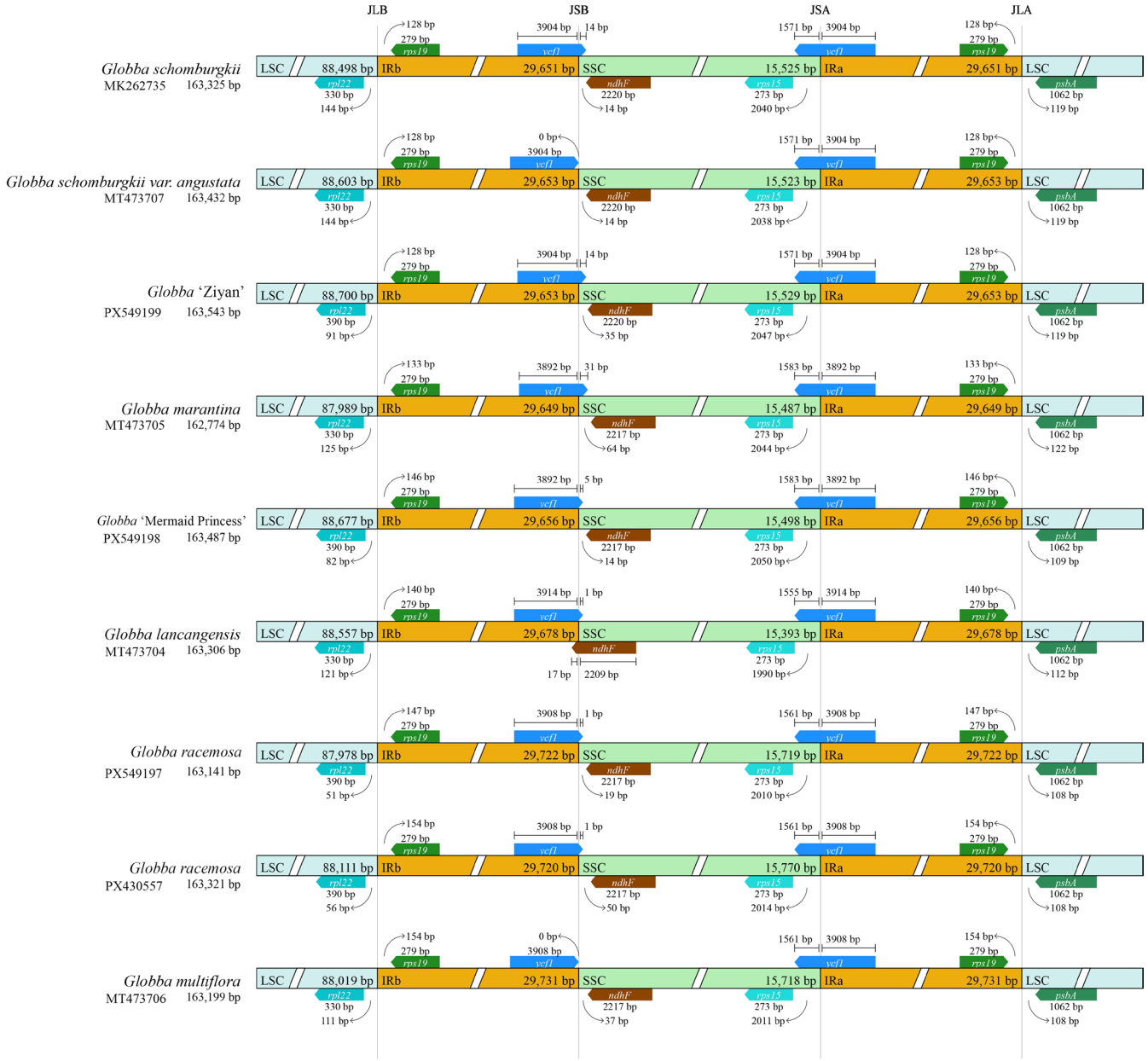


FIGURE 4 | Comparative analysis of the junctions between the inverted repeat (IR) regions and the two single-copy regions (large single-copy, LSC/small single-copy, SSC) in the nine *Globba* chloroplast genomes, which exhibited the typical angiosperm quadripartite structure. Colored boxes above or below the main line indicate adjacent border genes. The distance between the genes and boundaries is represented by the base lengths (bp). JLA, junction between LSC and inverted repeat (IRA). JLB, junction between LSC and IRB. JSA, junction between SSC and IRA. JSB, junction between SSC and IRB.

PR2-plot analysis revealed consistent compositional asymmetry at the third codon position across the 11 *Globba* chloroplast genomes. The protein-coding genes were predominantly clustered in higher-left quadrant (Q2, A3s < U3s and G3s > C3s), comprising 51.7%–53.5% of the total (Figure 8). The remaining genes were distributed across lower-left quadrant (Q3, 20.9%–23.0%), higher-right quadrant (Q1,17.2%–21.8%), and lower-right quadrant (Q4, 8.0%–8.1%). This skewed distribution, characterized by a strong bias toward G_3 over C_3 and U_3 over A_3, indicates a systematic mutational preference. The consistency of this pattern across all accessions suggests that the base composition at the third codon position in *Globba* chloroplast genomes is primarily shaped by mutation pressure rather than translational selection.

Neutrality-plot analysis revealed weak correlations between GC_3 and GC_{12} across the 11 *Globba* chloroplast genomes. All accessions showed low regression coefficients ($R^2 = 0.0268–0.0399$) and shallow slopes (0.1848–0.2300), with the highest slopes in *G. racemosa* (0.2300 and 0.2200) and the lowest in *G. marantina* (0.1848) (Figure 9). These consistently low R^2 values (all < 0.04) indicate that codon usage is shaped primarily by mutational pressure rather than selective constraints. Furthermore, the narrow range of slope values demonstrates the conservation of this evolutionary pattern across the genus. BUSTED analysis detected positive selection ($p \leq 0.05$) in two genes, including *infA* and *ndhB* (Table S3).

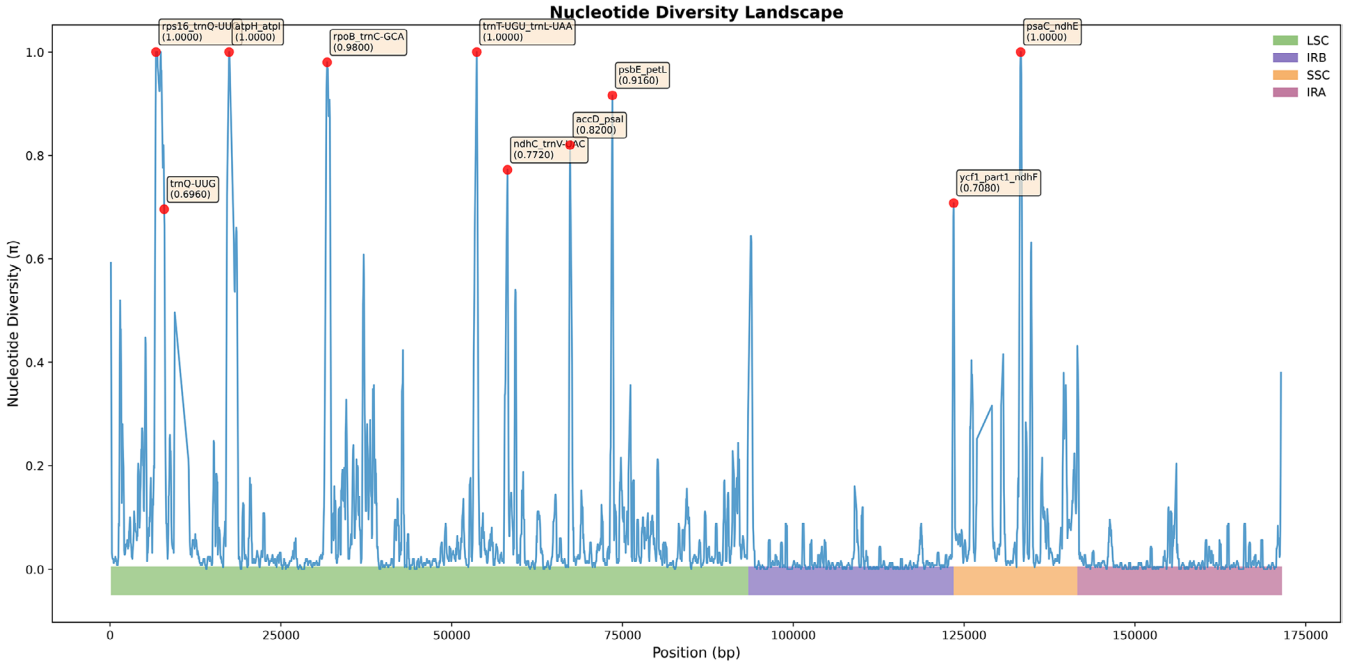


FIGURE 5 | Genes nucleotide diversity of 11 *Globba* species. Using *G. racemosa* (Enshi, China) PX549197 as the reference sequence.

3.6 | Phylogenetic Analysis of Chloroplast Genomes of *Globba* Species

A phylogenetic tree was constructed from the complete chloroplast genome sequences and 18 genes with high Pi values to elucidate the relationships among the 11 *Globba* species (Figure 10). The phylogenetic trees constructed based on the whole-genome sequences and the high-Pi genes were consistent in topology. However, the bootstrap support values differed, particularly for the two *G. racemosa* clades, which received 100% support based on the whole-genome data but only 55% support based on the high-Pi genes. All species were monophyletically grouped into two primary clades. One comprised *G. racemosa*, *G. multiflora*, *G. schomburgkii*, *G. schomburgkii* var. *angustata*, *G.* "Ziyan", *G. marantina*, and *G.* "Mermaid Princess". Within this clade, the two *G. racemosa* accessions formed a distinct subclade, confirming their close genetic relationship. The other consisted of *G. atrosanguinea*, *G. lancangensis*, and *G. variabilis*. Notably, the cultivated varieties *G.* "Ziyan" and *G.* "Mermaid Princess" were robustly placed within the same clade, showing closer phylogenetic affinity to *G. schomburgkii* and its variant than to other species in the clade. This strongly supported phylogeny provides a robust framework for species delineation and reconstructing the evolutionary history of the genus.

4 | Discussion

4.1 | Structural Conservation of Chloroplast Genomes

The analysis of chloroplast genomes in the genus *Globba* further confirms the high degree of structural conservation maintained (Tables 1 and 2; Figure 2) during the evolution of plant chloroplast genomes (Li et al. 2021). The total genome length, GC content, and number of genes showed minimal interspecific variation among the 11 species examined, consistent with observations in taxa such as *Pleurozia* (Bi et al. 2025) and *Rubia* (Li, Mo, and Wang 2025). While nine of the 11 samples exhibited the typical quadripartite structure, two species (*G. variabilis* and *G. atrosanguinea*) were found to lack fully complementary inverted repeat (IR) regions. This phenomenon has also been reported in genera such as *Rhododendron*, *Monotropeae*, *Medicago*, and *Erodium* (Shen et al. 2020).

4.2 | Gene Loss in Chloroplast Genomes

Analysis of the chloroplast genomes across *Globba* species, the total gene count exhibited slight variation, ranging from 132 to 134, due to differences in tRNA gene content. Specifically, the variation stemmed from the absence of either the *trnG-GCC* or *trnQ-UUG* gene. The *trnG-GCC* gene was missing in *G. multiflora* and *G. racemosa*, while *G. atrosanguinea* lacked the *trnQ-UUG* gene. This pattern of tRNA gene loss aligns with the phenomenon of structural plasticity observed in plant chloroplast genomes (Liu et al. 2019), where core photosynthetic and ribosomal genes remain highly conserved, while certain tRNA genes may undergo lineage-specific reduction. Such gene loss events have been documented in various plant and algae groups, including *Pleurozia* (An et al. 2023), *Chloropicon* and *Chloroparvula* (Turmel et al. 2019), suggesting that tRNA gene content may serve as a subtle but informative marker for tracking evolutionary divergence and phylogenetic relationships within closely related taxa.

4.3 | IR Boundary Variation and Expansion Patterns of *ndhF* and *ycf1* Genes

Expansion and contraction of IR boundaries are common phenomena in chloroplast genome evolution and are often

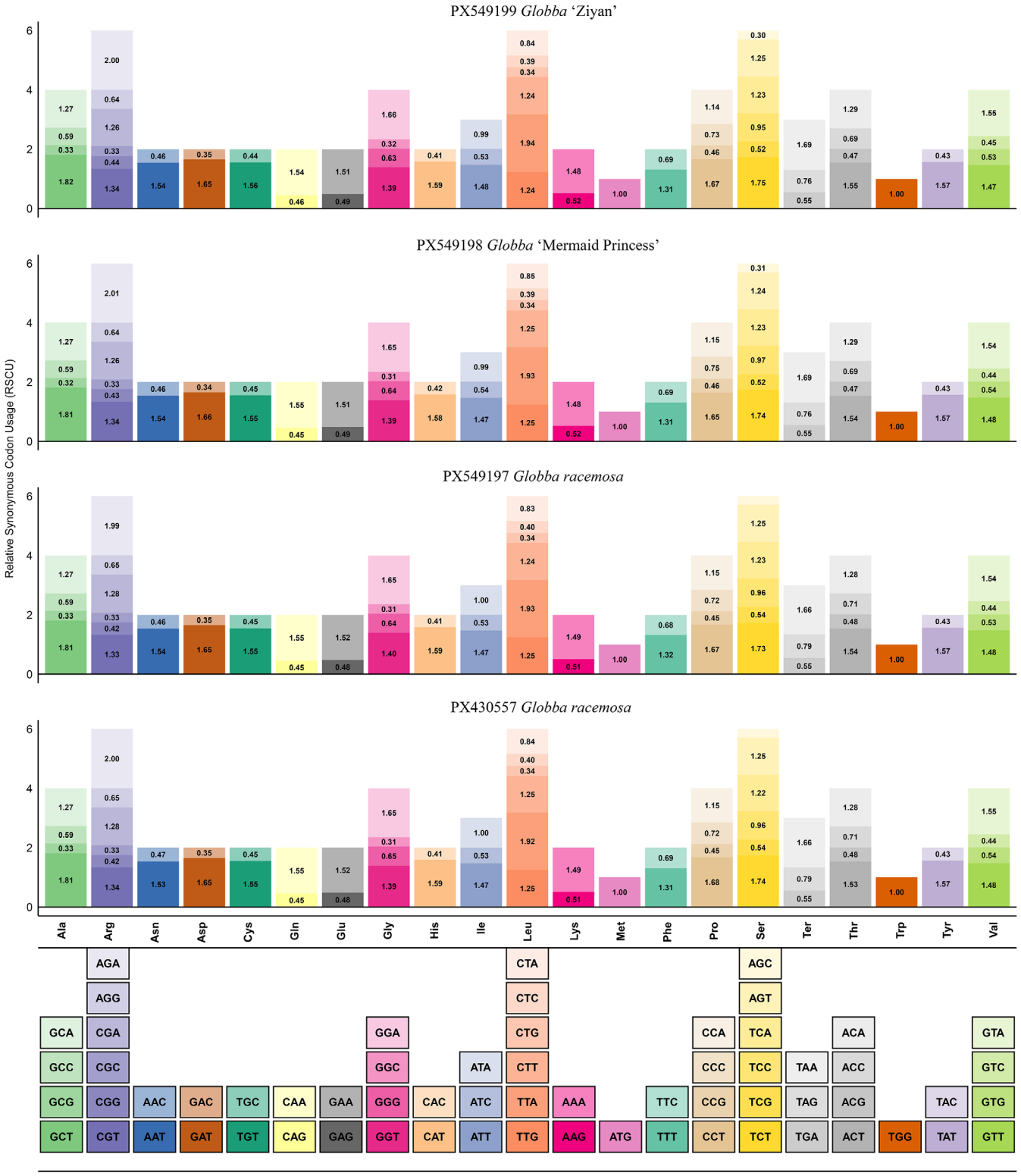


FIGURE 6 | RSCU-plot in the chloroplast genomes of 4 *Globba* species. Codons and corresponding amino acids are listed below the graph, above, and RSCU values on the color block.

accompanied by gene duplication, pseudogenization, or functional variation (Zhu et al. 2016). In the chloroplast genome of *G. lancangensis*, we observed that the *ndhF* gene uniquely spanned the IRb/SSC junction, extending 17 bp into the IRb region. Furthermore, the two copies of the *ycf1* gene exhibited strikingly asymmetric expansion across the IR/SSC boundaries: the copy in IRa extended 1555–1583 bp into the SSC, whereas the IRb copy extended only 1–31 bp (Figure 4). This expansion pattern was common among nine of the studied species, suggesting potential species-specific or evolutionary implications.

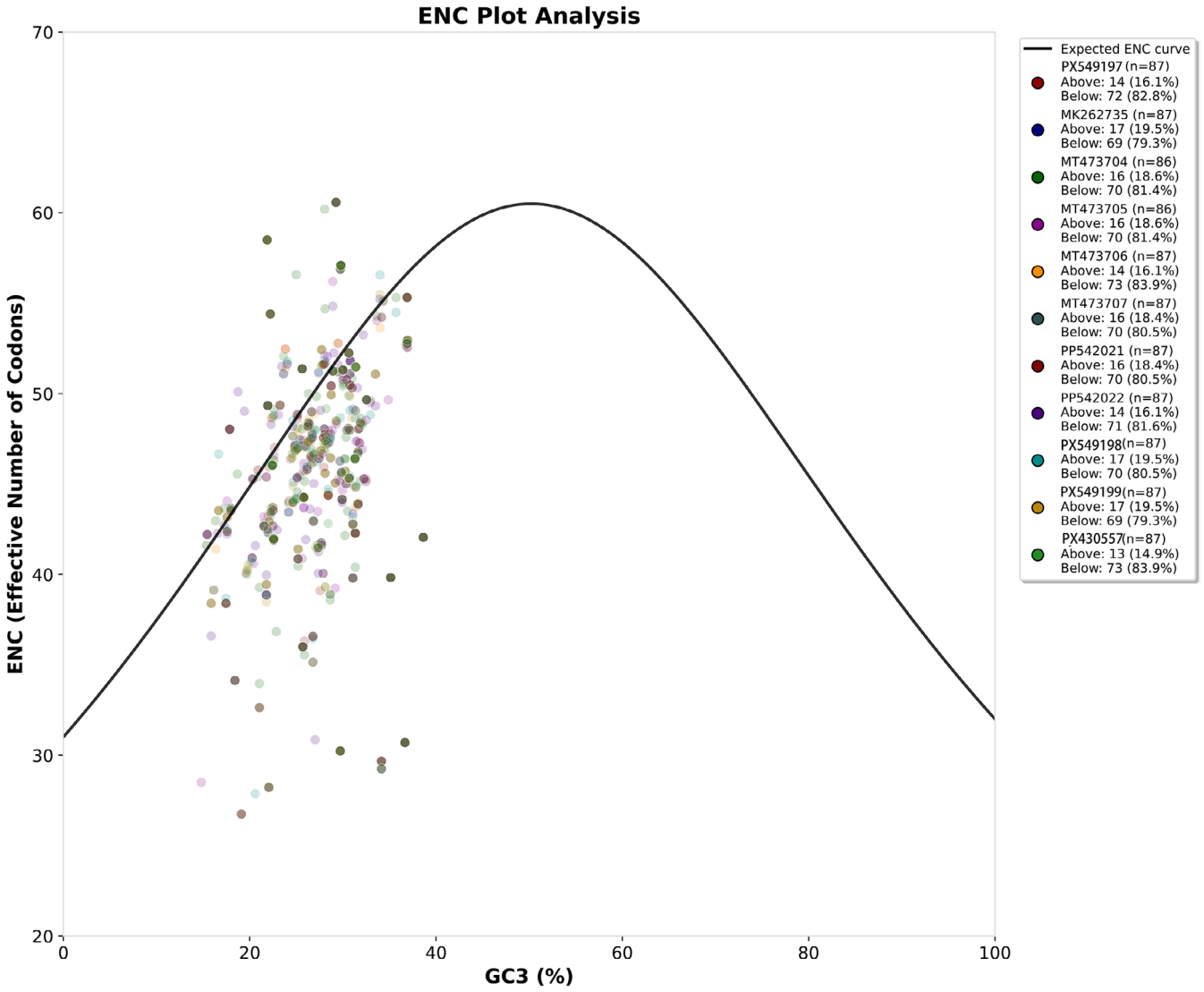


FIGURE 7 | ENC-GC3 plot of the chloroplast genomes of 11 *Globba* species.

The spanning of the IRb/SSC boundary by*c* has also been reported in other plant groups, possibly related to genomic recombination or repair mechanisms (Wang et al. 2008). The asymmetric expansion of *ycf1* across the IR/SSC junctions, especially the substantial extension from the IRa region, may reflect its specific role in maintaining genomic stability or environmental adaptation. Previous studies have highlighted that *ycf1* is one of the most variable regions in chloroplast genomes and has frequently been proposed as a potential DNA barcode (Dong et al. 2015; Neubig et al. 2008). The asymmetry in its expansion pattern could result from differential selection pressures or accumulated mutations, which warrant further functional investigation. Our study provides clues for future studies on species differentiation and adaptive evolution within this genus.

4.4 | Codon Usage Bias

Codon usage analysis of 11 *Globba* chloroplast genomes identified 17 optimal codons. The codon UUA (Leucine) showed the strongest bias, followed by ACU (Threonine) and CGU (Arginine). These optimal codons represent several amino acids, including leucine, threonine, arginine, serine, and glycine, with a clear predominance of A/U-ending codons (Table S1). This preference for A/U-ending codons aligns with the common pattern observed in most angiosperm chloroplast genomes and is often attributed to a balance between mutation pressure and natural selection (Gustafsson et al. 2004; Abdullah et al. 2021; Song et al. 2024).

Analysis of the ENC revealed two distinct groups among the protein-coding genes: a high-expression group (95 genes, ENC: 26.75–38.86) and a low-expression group (95 genes, ENC: 52.78–60.58) (Figure 7). Only the *psaC* gene had an ENC value greater than 60, while *rpl32* and *psbJ* had values below 30, suggesting stronger translational selection or functional constraints for these latter genes (Sharp and Li 1987). A comparative analysis among species showed that *G. schomburgkii*, *G. "Mermaid Princess"*, and *G. "Ziyan"* had a higher proportion of genes (over 19%) plotting above the expected ENC curve. In contrast, *G.*

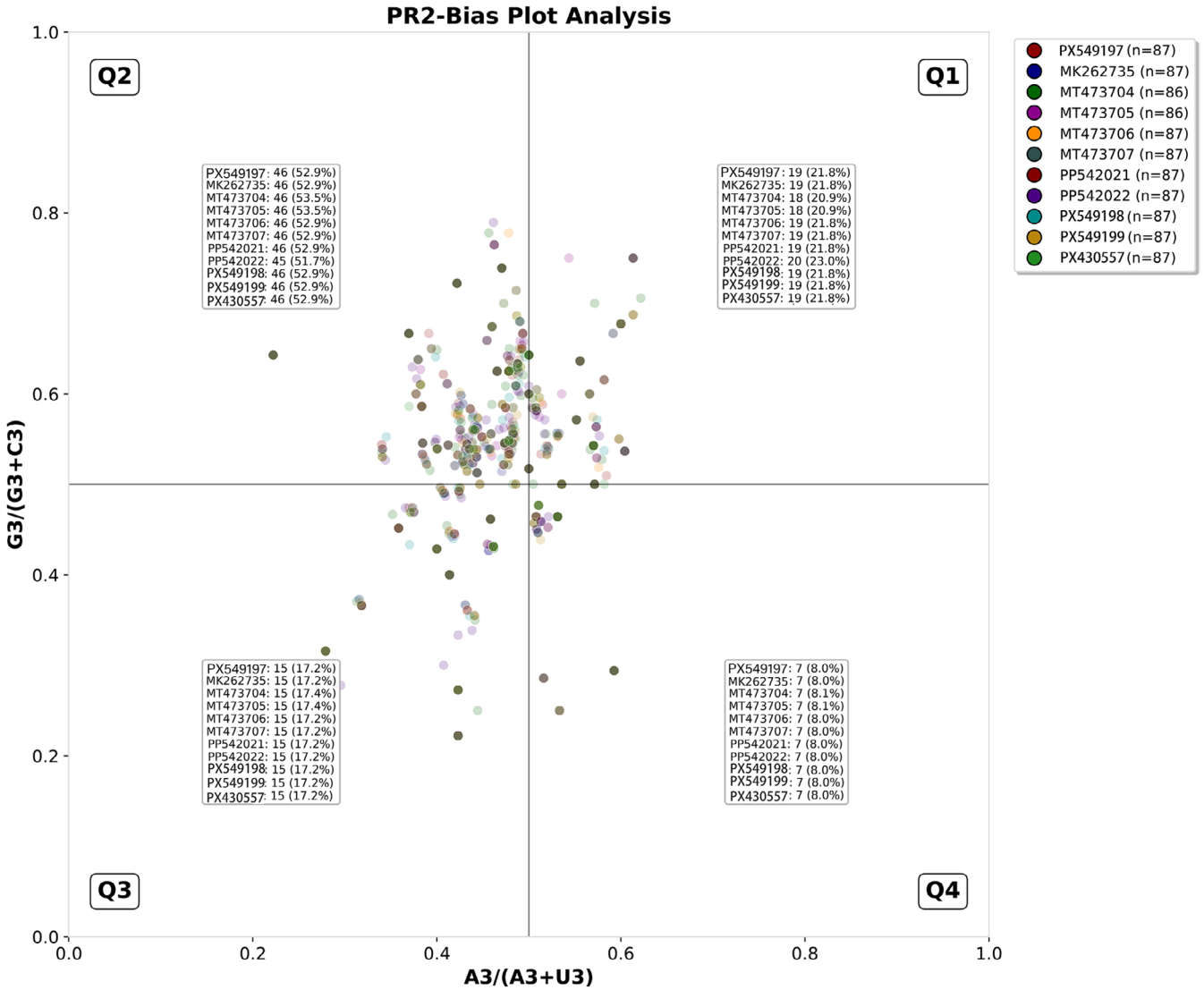


FIGURE 8 | PR2-plot analysis of the chloroplast genomes of 11 *Globba* species.

racemosa, *G. multiflora*, and *G. variabilis* displayed a lower proportion (under 17%). Notably, a clear difference was observed between the two *G. racemosa* samples from Quzhou and Enshi, with 16.1% and 14.9% of genes above the curve, respectively, suggesting potential intraspecific variation in codon usage patterns.

4.5 | Analysis of Selection Pressure

To elucidate the evolutionary forces shaping codon usage in *Globba* chloroplast genomes, PR2-plot and neutrality plot analyses were conducted. The PR2-plot revealed consistent compositional asymmetry at the third codon position across all 11 species. Protein-coding genes were predominantly clustered in the upper-left quadrant (Q2, A3s < U3s and G3s > C3s), accounting for 51.7% to 53.5% of the total (Figure 8). This systematic bias toward G3 over C3 and U3 over A3 indicates that mutation pressure, rather than translational selection, is the primary determinant of base composition at the synonymous third codon position in *Globba* chloroplast genomes (Sueoka 1988).

Neutrality plot analysis further supports this conclusion. The correlation between GC_3 and GC_{12} was weak across all samples, with low regression coefficients ($R^2 = 0.0268–0.0399$) and flat slopes (0.1848–0.2300). *G. racemosa* exhibited the highest slope (0.2300 and 0.2200 for the two samples), while *G. marantina* showed the lowest (0.1848) (Figure 9). The consistently low R^2 values (all < 0.04) suggest that mutation pressure is the dominant force influencing codon usage patterns in this genus (Wright 1990; Li, Liu, et al. 2025). The narrow range of slope values indicates a conserved evolutionary mode within *Globba*. Notably, a positive selection signal ($p \leq 0.05$) was detected in two genes, *infA* and *ndhB*, using BUSTED analysis (Table S3), implying that these genes may have undergone adaptive evolution, possibly in response to specific environmental or functional demands (Murrell et al. 2015).

4.6 | Phylogenetic Relationships

A phylogenetic tree, constructed from complete chloroplast genome sequences, robustly supported the monophyly of the

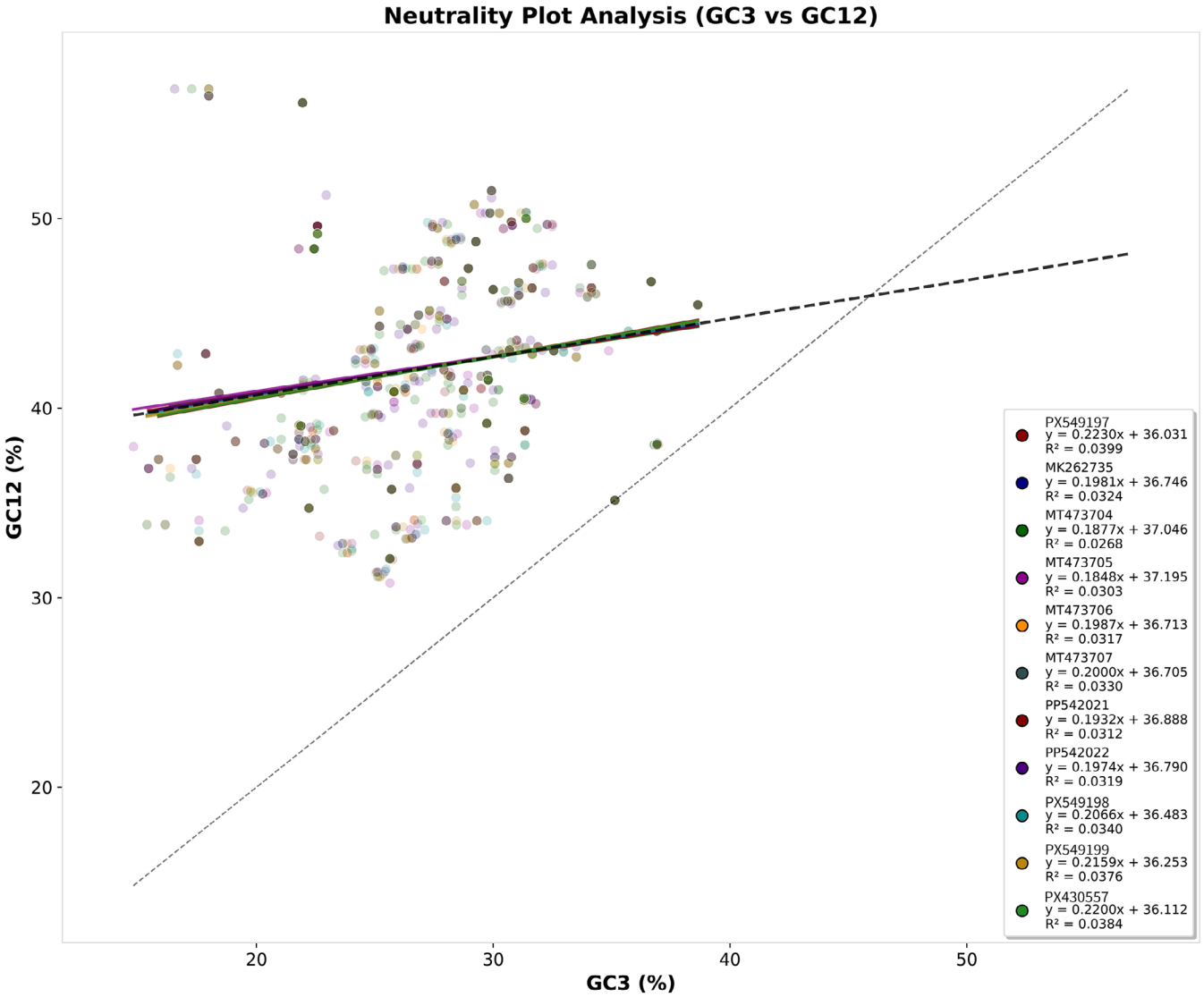


FIGURE 9 | Neutral-plot analysis of the chloroplast genomes of 11 *Globba* species.

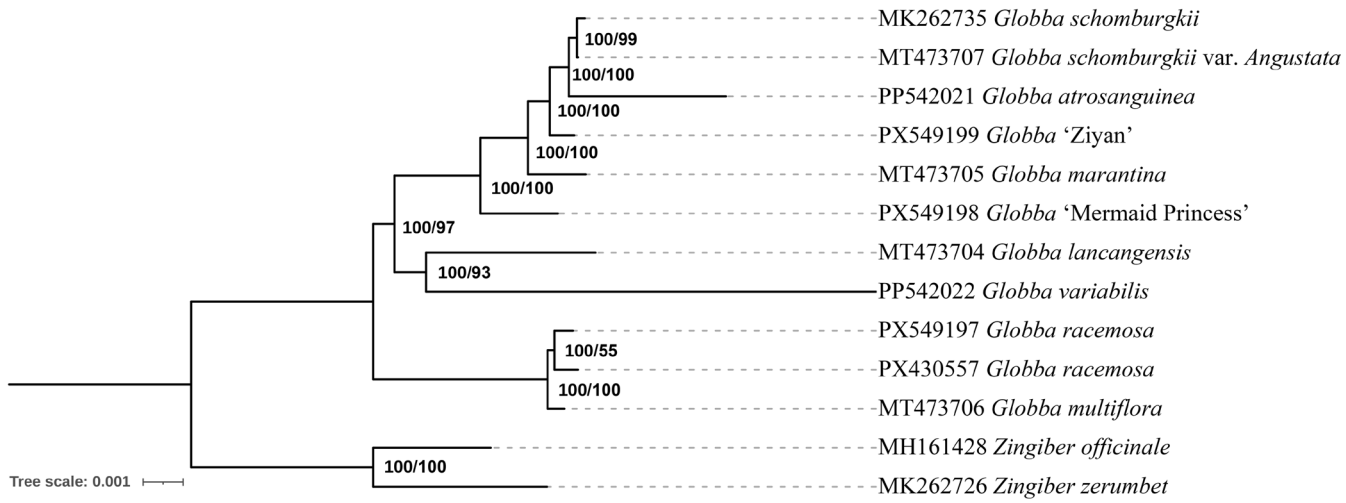


FIGURE 10 | Phylogenetic tree of *Globba* species based on complete chloroplast genomes and 18 genes with high Pi values. (Support values are shown as follows: Those to the left of the slash (/) were obtained from the whole-genome dataset, and those to the right were obtained from the high-Pi genes.)

studied *Globba* species and resolved them into two major clades (Figure 10). The two *G. racemosa* accessions formed a well-supported subclade, indicating their close genetic relationship. The two cultivars, *G.* "Ziyan" and *G.* "Mermaid Princess", were deeply embedded within Clade I, showing closer phylogenetic ties to *G. schomburgkii* and its variety than to other species in this lineage. This well-supported placement provides molecular evidence for their breeding origins and aids in cultivar identification. The overall topology is largely congruent with previous studies based on morphology or a few gene fragments but offers substantially higher resolution and support due to the use of whole chloroplast genomes (Jabbour and Renner 2012). The robust phylogenetic framework established here provides a solid foundation for species delimitation and taxonomic revision within *Globba* and offers crucial evolutionary context for future investigations into its biogeographic dispersal, hybridization events, and adaptive radiation (Wen et al. 2018).

5 | Conclusion

This study reports the complete chloroplast genomes of *Globba racemosa*, *G.* "Mermaid Princess", and *G.* "Ziyan", and characterizes the chloroplast genome features, codon usage bias, and evolution of 11 *Globba* species. The total number of genes ranged from 132 to 134, which is attributed to variations in tRNA gene counts (37–39). The chloroplast genomes exhibit the typical quadripartite structure of *Globba* plants, showing high conservation in size and gene content, as well as notably asymmetric expansion across the IR/SSC boundaries in nine species. In the remaining two species, however, no long inverted repeat sequences were detected; instead, numerous short inverted repeats were observed, a pattern inconsistent with the typical characteristics of chloroplast genomes. ENC-Plot, PR2-plot, Neutral-plot, and BUSTED analysis suggest that mutation pressure is the dominant force influencing codon usage patterns in the *Globba* genus. Phylogenetic analyses based on complete chloroplast genome sequences robustly resolved both intergeneric and intrageneric relationships within the *Globba* genus. These newly generated genomic resources provide a foundation for further studies in taxonomy, systematics, and conservation biology of these medicinally significant species.

Author Contributions

Wenhui Liao: writing – original draft (equal). **Xianlin Rong:** conceptualization (equal), resources (equal). **Chunhai Wei:** investigation (equal), visualization (equal). **Mingli Zhang:** investigation (equal), resources (equal). **Jianshuang Shen:** conceptualization (equal), data curation (equal), funding acquisition (equal), investigation (equal), methodology (equal), writing – original draft (equal), writing – review and editing (equal).

Acknowledgments

The authors have nothing to report.

Funding

This work was supported by the National Natural Science Foundation of China [32201608]; the Research Foundation for Advanced Talents of Hangzhou Vocational & Technical College [RCXY202303]; 2024 Annual Domestic Visiting Engineer "School-Enterprise Cooperation Project" for Higher Education Institutions (FG2024088).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The genome sequence data that support the findings of this study are openly available in GenBank of the NCBI, the associated BioProject, SRA, and Bio-Sample numbers of Globba racemosa Smith collected from Quzhou, China, are PRJNA1335541, SRR35647342 and SAMN51991279; Globba racemosa Smith collected from Enshi, China, are PRJNA1366750, SRR36106138 and SAMN53322884; G. 'Mermaid Princess' are PRJNA1366750, SRR36106137 and SAMN53322885; G. 'Ziyan' are PRJNA1366750, SRR36106136 and SAMN53322886.

References

Abdullah, F., A. Mehmood, A. Rahim, P. Heidari, I. Ahmed, and P. Poczai. 2021. "Comparative Plastome Analysis of *Blumea*, With Implications for Genome Evolution and Phylogeny of Asteroideae." *Ecology and Evolution* 11, no. 12: 7810–7826. https://doi.org/10.1002/ece3.7614.

An, S. M., B. Y. Kim, H. Kang, et al. 2023. "The Complete Chloroplast Genome Sequences of Six *Hylotelephium* Species: Comparative Genomic Analysis and Phylogenetic Relationships." *PLoS One* 10, no. 18: e0292056. https://doi.org/10.1371/journal.pone.0292056.

Beier, S., T. Thiel, T. Münch, U. Scholz, and M. Mascher. 2017. "MISA-Web: A Web Server for Microsatellite Prediction." *Bioinformatics* 33, no. 16: 2583–2585. https://doi.org/10.1093/bioinformatics/btx198.

Bi, S., Q. Liu, J. W. Hao, X. Z. Cai, D. Gao, and L. N. Zhang. 2025. "Complete Chloroplast Genomes of Three Pleurozia Species and Comparative Analyses With *P. purpurea*: Codon Usage Bias and Phylogeny." *Frontiers in Plant Science* 16: 1599291. https://doi.org/10.3389/fpls.2025.1599291.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. "Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor." *Bioinformatics* 34, no. 17: i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Comeron, J. M., and M. Aguadé. 1998. "An Evaluation of Measures of Synonymous Codon Usage Bias." *Journal of Molecular Evolution* 47, no. 3: 268–274. https://doi.org/10.1007/pl00006384.

Dong, W., C. Xu, C. Li, et al. 2015. "*ycf1*, the Most Promising Plastid DNA Barcode of Land Plants." *Scientific Reports* 5: 8348. https://doi.org/10.1038/srep08348.

Doungchawee, J., C. Kulsing, N. Suekaew, et al. 2019. "Volatile Chemical Composition, Antibacterial and Anti- Fungal Activities of Extracts From Different Parts of *Globba schomburgkii* Hook. F." *Chemistry and Biodiversity* 16, no. 5: e1900057. https://doi.org/10.1002/cbdv.201900057.

Doyle, J. J. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue." *Phytochemical Bulletin* 19: 11–15.

Gustafsson, C., S. Govindarajan, and J. Minshull. 2004. "Codon Bias and Heterologous Protein Expression." *Trends in Biotechnology* 22: 346–353. https://doi.org/10.1016/j.tibtech.2004.04.006.

Jabbour, F., and S. S. Renner. 2012. "A Phylogeny of Delphinieae (Ranunculaceae) Shows That *Aconitum* Is Nested Within *Delphinium* and That Late Miocene Transitions to Long Life Cycles in the Himalayas and Southwest China Coincide With Bursts in Diversification." *Molecular Phylogenetics and Evolution* 62, no. 3: 928–942. https://doi.org/10.1016/j.ympev.2011.12.005.

Johnson, M., I. Zaretskaya, Y. Raytselis, Y. Merezhuk, S. McGinnis, and T. L. Madden. 2008. "NCBI BLAST: A Better Web Interface." *Nucleic*

Acids Research 36, no. suppl_2: W5–W9. https://doi.org/10.1093/nar/gkn201.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780. https://doi.org/10.1093/molbev/mst010.

Kearse, M., R. Moir, A. Wilson, et al. 2012. "Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data." *Bioinformatics* 28, no. 12: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kosakovsky Pond, S. L., A. F. Y. Poon, R. Velazquez, et al. 2020. "HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies." *Molecular Biology and Evolution* 37, no. 1: 295–299. https://doi.org/10.1093/molbev/msz197.

Kress, W. J., L. M. Prince, and K. J. Williams. 2002. "The Phylogeny and A New Classification of the Gingers (Zingiberaceae): Evidence From Molecular Data." *American Journal of Botany* 2002: 1682–1696. https://doi.org/10.3732/ajb.89.10.1682.

Kuncharoen, N., P. Pittayakhajonwut, and S. Tanasupawat. 2018. "*Micromonospora globbae* sp. Nov., An Endophytic Actinomycete Isolated From Roots of *Globba winitii* C. H. Wright." *International Journal of Systematic and Evolutionary Microbiology* 68, no. 4: 1073–1077. https://doi.org/10.1099/ijsem.0.002625.

Li, D. M., J. Li, Y. C. Wang, Y. C. Xu, and G. F. Zhu. 2021. "Molecular Evolution of Chloroplast Genomes in Subfamily Zingiberoideae (Zingiberaceae)." *BMC Plant Biology* 21: 558. https://doi.org/10.1186/s12870-021-03315-9.

Li, H. 2021. "New Strategies to Improve Minimap2 Alignment Accuracy." *Bioinformatics* 37: 4572–4574. https://doi.org/10.1093/bioinformatics/btab705.

Li, H., Q. Guo, L. Xu, H. Gao, L. Liu, and X. Zhou. 2023. "CPJSdraw: Analysis and Visualization of Junction Sites of Chloroplast Genomes." *PeerJ* 11: e15326. https://doi.org/10.7717/peerj.15326.

Li, H. D., C. Zhao, D. P. Wang, et al. 2017. "GC-MS Analysis of Volatile Oils From Rhizome, Stem, and Leaf of *Globba racemosa*." *Journal of Chinese Medicinal Materials* 40, no. 6: 1351–1354. https://doi.org/10.13863/j.issn1001-4454.2017.06.025.

Li, X., X. Mo, and D. Wang. 2025. "Phylogeny and Evolutionary Dynamics of the *Rubia* Genus Based on the Chloroplast Genome of *Rubia tibetica*." *Scientific Reports* 15: 14370. https://doi.org/10.1038/s41598-025-96665-5.

Li, X. J., L. E. Liu, Q. D. Ren, et al. 2025. "Analysis of Synonymous Codon Usage Bias in the Chloroplast Genome of Five Caragana." *BMC Plant Biology* 25: 322. https://doi.org/10.1186/s12870-025-06351-x.

Liu, E., C. Yang, J. Liu, et al. 2019. "Comparative Analysis of Complete Chloroplast Genome Sequences of Four Major *Amorphophallus* Species." *Scientific Reports* 9, no. 1: 809. https://doi.org/10.1038/s41598-018-37456-z.

Liu, S. Y., Y. Ni, J. L. Li, et al. 2023. "CPGView: A Package for Visualizing Detailed Chloroplast Genome Structures." *Molecular Ecology Resources* 23: 694–704. https://doi.org/10.1111/1755-0998.

Lowe, T. M., and S. R. Edd. 1997. "tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence." *Nucleic Acids Research* 25, no. 5: 955–964. https://doi.org/10.1093/nar/25.5.955.

Matsumoto, A., and Y. Takahashi. 2019. "Endophytic Actinomycetes: Promising Source of Novel Bioactive Compounds." *Journal of Antibiotics (Tokyo)* 70, no. 5: 514–519. https://doi.org/10.1038/ja.2017.20.

Menon, A. N., and M. Dan. 2009. "Chemical Composition of Essential Oils of *Globba* Species From South India." *Journal of Essential Oil Research* 21, no. 1: 59–60. https://doi.org/10.1080/10412905.2009.9700110.

Murrell, B., S. Weaver, M. D. Smith, et al. 2015. "Gene-Wide Identification of Episodic Selection." *Molecular Biology and Evolution* 32, no. 5: 1365–1371. https://doi.org/10.1093/molbev/msv035.

Naznin, S., M. Ferdous, M. S. Uddin, F. Yeasmin, A. S. Daula, and M. A. Basher. 2019. "Phytochemical Investigation and Evaluation of Antioxidant Activities of Methanolic Extract of *Globba marantina* Leaves." *Journal of Medicinal Plants Studies* 7, no. 6: 147–153.

Neubig, K. M., W. M. Whitten, B. S. Carlsward, et al. 2008. "Phylogenetic Utility of *ycf1* in Orchids: A Plastid Gene More Variable Than *matK*." *Plant Systematics and Evolution* 277: 75–84. https://doi.org/10.1007/s00606-008-0105-0.

Qian, R., D. Xu, Y. He, and H. Zhang. 2013. "Feasibility Analysis and Countermeasures for Establishing the Qianligang Nature Reserve in Qujiang District, Zhejiang Province." *East China Forest Management* 27, no. 4: 60–62. https://wenku.baidu.com/view/ecca62826f1aff00bfd51e34?_wkts_=1773304877139.

Sangvirotjanapat, S., J. Denduangboriphant, and M. F. Newman. 2019. "A Taxonomic Revision of *Globba* Subsect. *Nudae* (Zingiberaceae)." *European Journal of Taxonomy* 505: 1–7. https://doi.org/10.5852/EJT.2019.503.

Santra, H. K., and D. Bera. 2025. "Bioactivities of Secondary Metabolites of Two Actinomycetes *Streptomyces parvulus* GloL3 and *Streptomyces lienomycini* SK5, Endophytes of *Globba marantina* L. and *Selaginella kraussiana*." *Brazilian Journal of Microbiology* 56: 779–795. https://doi.org/10.1007/s42770-025-01615-y.

Sharp, P. M., and W. H. Li. 1986. "An Evolutionary Perspective on Synonymous Codon Usage in Unicellular Organisms." *Journal of Molecular Evolution* 24: 28–38. https://doi.org/10.1007/BF02099948.

Sharp, P. M., and W. H. Li. 1987. "The Codon Adaptation Index—A Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications." *Nucleic Acids Research* 15, no. 3: 1281–1295. https://doi.org/10.1093/nar/15.3.1281.

Shen, J., X. Li, X. Zhu, and X, Huang, and S. Jin. 2020. "The Complete Plastid Genome of *Rhododendron pulchrum* and Comparative Genetic Analysis of *Ericaceae* Species." *Forests* 11, no. 2: 158. https://doi.org/10.3390/f11020158.

Shen, J., W. Liao, X. Rong, C. Wei, and M. Zhang. 2026. "Characterization of the Complete Chloroplast Genomes of *Globba racemosa* Smith in the Family Zingiberaceae." *Mitochondrial DNA Part B Resources* 11, no. 2: 237–240. https://doi.org/10.1080/23802359.2025.2611470.

Song, C., J. Zhu, and H. Li. 2024. "Complete Chloroplast Genomes of Eight *Delphinium* Taxa (Ranunculaceae) Endemic to Xinjiang, China: Insights Into Genome Structure, Comparative Analysis, and Phylogenetic Relationships." *BMC Plant Biology* 24, no. 1: 1–22. https://doi.org/10.1186/s12870-024-05279-y.

Sueoka, N. 1988. "Directional Mutation Pressure and Neutral Molecular Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 85: 2653–2657. https://doi.org/10.1073/pnas.85.8.2653.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner. 2003. "Exploiting EST Databases for the Development and Characterization of Gene-Derived SSR-Markers in Barley (*Hordeum vulgare* L.)." *Theoretical and Applied Genetics* 106, no. 3: 411–422. https://doi.org/10.1007/s00122-002-1031-0.

Tillich, M., P. Lehwark, T. Pellizzer, et al. 2017. "GeSeq–Versatile and Accurate Annotation of Organelle Genomes." *Nucleic Acids Research* 45, no. W1: W6–W11. https://doi.org/10.1093/nar/gkx391.

Turmel, M., A. Lopes Dos Santos, C. Otis, R. Sergerie, and C. Lemieux. 2019. "Tracing the Evolution of the Plastome and Mitogenome in the

Chloropicophyceae Uncovered Convergent tRNA Gene Losses and a Variant Plastid Genetic Code." *Genome Biology and Evolution* 11, no. 4: 1275–1292. https://doi.org/10.1093/gbe/evz074.

Wang, R. J., C. L. Cheng, C. C. Chang, C. L. Wu, T. M. Su, and S. M. Chaw. 2008. "Dynamics and Evolution of the Inverted Repeat-Large Single Copy Junctions in the Chloroplast Genomes of Monocots." *BMC Evolutionary Biology* 8, no. 1: 36. https://doi.org/10.1186/1471-2148-8-36.

Wen, J., A. J. Harris, Y. Kalburgi, et al. 2018. "Chloroplast Phylogenomics of the New World Grape Species (Vitis, Vitaceae)." *Journal of Systematics and Evolution* 56, no. 4: 297–308. https://doi.org/10.1111/jse.12447.

Williams, K., W. J. Kress, and P. S. Manos. 2004. "The Phylogeny, Evolution, and Classification of the Genus *Globba* and Tribe Globbeae (Zingiberaceae): Appendages Do Matter." *American Journal of Botany* 91, no. 1: 100–114. https://doi.org/10.3732/ajb.91.1.100.

Wright, F. 1990. "The 'Effective Number of Codons' Used in a Gene." *Gene* 87, no. 1: 23–29. https://doi.org/10.1016/0378-1119(90)90491-9.

Zhu, A., W. Guo, S. Gupta, W. Fan, and J. P. Mower. 2016. "Evolutionary Dynamics of the Plastid Inverted Repeat: The Effects of Expansion, Contraction, and Loss on Substitution Rates." *New Phytologist* 209: 1747–1756. https://doi.org/10.1111/nph.13743.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Figure S1:** Schematic map of the cis-splicing and trans-splicing genes in the chloroplast genomes of three *Globba* species. (A) *G. racemosa* Smith (Enshi, China), PX549197; (B) *G.* "Mermaid Princess," PX549198; (C) *G.* "Ziyan," PX549199. **Figure S2:** RSCU-plot in the chloroplast genomes of 11 *Globba* species. **Table S1:** GC content and ENC in the chloroplast genomes of 11 Globba species. **Table S2:** RSCU in the chloroplast genomes of 11 Globba species. **Table S3:** Results of the gene analysis by BUSTED (Omega value calculation).