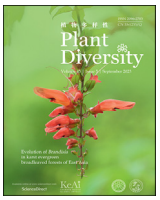
Plant Diversity 45 (2023) 523–534

Contents lists available at ScienceDirect

Plant Diversity

journal homepage: http://www.keaipublishing.com/en/journals/plant-diversity/
http://journal.kib.ac.cn



Research paper

Plastid phylogenomics and species discrimination in the "Chinese" clade of *Roscoea* (Zingiberaceae)

Hai-Su Hu [a, b], Jiu-Yang Mao [b], Xue Wang [b], Yu-Ze Liang [b], Bei Jiang [b, c], De-Quan Zhang [a, b, c, *]

[a] School of Life Science and Engineering, Southwest University of Science and Technology, Mianyang 621010, Sichuan, China
[b] College of Pharmacy, Dali University, Dali 671000, Yunnan, China
[c] Yunnan Key Laboratory of Screening and Research on Anti-pathogenic Plant Resources from Western Yunnan (Cultivation), Dali 671000, Yunnan, China

ARTICLE INFO

Article history:
Received 21 November 2022
Received in revised form
15 March 2023
Accepted 30 March 2023
Available online 6 April 2023

Keywords:
Medicinal plant
Chloroplast genome
Molecular phylogeny
DNA barcoding
DNA sequencing
Species identification

ABSTRACT

Roscoea is an alpine or subalpine genus from the pan-tropical family Zingiberaceae, which consists of two disjunct groups in geography, namely the "Chinese" clade and the "Himalayan" clade. Despite extensive research on the genus, *Roscoea* species remain poorly defined and relationships between these species are not well resolved. In this study, we used plastid genomes of nine species and one variety to resolve phylogenetic relationships within the "Chinese" clade of *Roscoea* and as DNA super barcodes for species discrimination. We found that *Roscoea* plastid genomes ranged in length from 163,063 to 163,796 bp, and encoded 113 genes, including 79 protein-coding genes, 30 tRNA genes, four rRNA genes. In addition, expansion and contraction of the IR regions showed obvious infraspecific conservatism and interspecific differentiation. Plastid phylogenomics revealed that species belonging to the "Chinese" clade of *Roscoea* can be divided into four distinct subclades. Furthermore, our analysis supported the independence of *R. cautleoides* var. *pubescens*, the recovery of *Roscoea pubescens* Z.Y. Zhu, and a close relationship between *R. humeana* and *R. cautloides*. When we used the plastid genome as a super barcode, we found that it possessed strong discriminatory power (90%) with high support values. Intergenic regions provided similar resolution, which was much better than that of protein-coding regions, hypervariable regions, and DNA universal barcodes. However, plastid genomes could not completely resolve *Roscoea* phylogeny or definitively discriminate species. These limitations are likely related to the complex history of *Roscoea* speciation, poorly defined species within the genus, and the maternal inheritance of plastid genomes.

Copyright © 2023 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Roscoea Smith. is an economically important alpine or subalpine genus from the pan-tropical family Zingiberaceae. The genus includes approximately 20 species worldwide, although most are distributed in the Sino-Himalayan region (Cowley, 2007). A total of 13 *Roscoea* species grow in China, including eight endemics, which primarily grow at high elevations in the subtropical and warm temperate zones of southwest China, including Yunnan, Sichuan, and Tibet (Xizang) (Wu and Larsen, 2000; Cowley, 2007; Zhao et al., 2016a). *Roscoea* species are used in both traditional Chinese medicine (Luo et al., 2008; Sahu et al., 2010; Srivastava et al., 2015; Rawat et al., 2018) and are harvested for horticultural purposes (Misra et al., 2015). In addition, *Roscoea* is a special genus within Zingiberaceae owing to its distribution at high elevations (Zhao et al., 2016a) and thus merits greater attention.

The *Roscoea* species have been divided into two distinct groups, namely a "Chinese" clade and a "Himalayan" clade (Ngamriabsakul et al., 2000; Cowley, 2007; Zhao et al., 2017). Although these two clades have been supported by previous studies, relationships between species within the genus have not been well resolved (Zhao et al., 2017). Researchers have suggested that the poor resolution within the genus may be attributed to morphological complexity (Zhao et al., 2017), recent species divergence (Zhao et al., 2016a, 2016b, 2021), natural hybridization, and/or introgressions (Du et al., 2012; Zhao et al., 2017). One additional limitation of previous

*Corresponding author. School of Life Science and Engineering, Southwest University of Science and Technology, Mianyang, 621010, Sichuan, China
E-mail address: zhangdeq2008@126.com (D.-Q. Zhang).
Peer review under responsibility of Editorial Office of Plant Diversity.

https://doi.org/10.1016/j.pld.2023.03.012
2468-2659/Copyright © 2023 Kunming Institute of Botany, Chinese Academy of Sciences. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

molecular studies is that they were based single-DNA fragments, such as ITS, *trnH-psbA*, *trnL-F* (Zhao et al., 2017), ITS (Ngamriababsakul et al., 2000), or SNPs (Zhao et al., 2021). Plastid genome sequences may offer an approach to reveal phylogenetic relationships within *Roscoea*. Several studies have used plastid genomes to resolve recalcitrant plant lineages (Wei et al., 2021; Escobari et al., 2021; Kleinwee et al., 2022; Xu et al., 2022). Plastid genomes are relatively easy to assemble and annotate now, and they can provide valuable resources to assess inter-specific relationships, especially within unresolved low taxonomic levels (Li et al., 2021). However, to our knowledge, no studies have attempted to resolve phylogenetic relationships within *Roscoea* by using plastid genome sequences.

Some *Roscoea* species are extremely difficult to identify, such as *R. tibetica*. These plants have unstable morphological traits at different growth phases and the morphological characters of many specimens are distorted. In our previous study, the universal barcodes were adopted to identify species in *Roscoea* and showed high success rates of ITS + *trnH-psbA* but low reliability in some species due to the limited informative sites (Zhang et al., 2014). One possible solution to this problem is using the plastid genome as a super barcode. Plastid genomes have been widely used to discriminate species that have recently diverged and/or complicated genera, and may be superior to many universal DNA barcodes (Li et al., 2015; Ji et al., 2019; Fu et al., 2019, 2022; Chen et al., 2022; Zhang et al., 2022a).

In this study, we used plastid genomes to clarify the phylogenetic relationships within the “Chinese” clade of *Roscoea*, and also evaluated plastid genome sequences as super barcodes for these same species. For these purposes, we used NGS technology to sequence and then annotate the plastid genomes of nine species and one variety within the “Chinese” clade of *Roscoea*. This work will be beneficial to future studies on the phylogeny, taxonomy, and conservation of *Roscoea*.

2. Materials and methods

2.1. Material sampling

We sampled fresh leaves from 28 healthy and mature individuals of nine species and one variety of the “Chinese” clade of *Roscoea*. Most samples were collected from their natural habitats in the Hengduan Mts. and adjacent regions (Fig. 1; Table S1). To avoid sampling individuals from the same female parent, we collected two to four individuals separated by > 30 m for each species. Samples were immediately dried using silica gel.

Plants at flowering stage were excavated at each location, and used as voucher specimens for the species. All specimens were taxonomically identified according to the *Flora of China* (Wu and Larsen, 2000) and the Genus *Roscoea* (Cowley, 2007). Our previous work (Zhang et al., 2014, 2015) indicated that *R. cautleoides* var. *pubescens* should be recovered as *R. pubescens* Z.Y. Zhu; however, here we still adopt the taxonomy in *Flora of China* (Wu and Larsen, 2000). Voucher specimens were deposited at the Herbarium of Medicinal Plants and Crude Drugs of the college of Pharmacy, Dali University (Table S1).

2.2. DNA extraction and next-generation sequencing (NGS)

Total genomic DNA was extracted from the dried leaf material using a modified CTAB method (Doyle, 1987; Yang et al., 2014). A total amount of 0.2 μg of DNA per sample was used for the DNA library preparations. Genomic DNA samples were fragmented to a size of 350 bp using sonication. Thereafter, the DNA fragments were end-polished, A-tailed, ligated with the full-length adapter, and amplified using PCR. The product was separated using electrophoresis on a 1.0% agarose gel, purified using a gel extraction kit, and used for construction of the DNA library (Yang et al., 2014). After the library passed quality inspection, the DNA was sequenced based on the NGS technique implemented on the Illumina HiSeq 2500 platform with paired-end sequencing (2 × 150 bp). The entire high-throughput sequencing was conducted by Novogene Bioinformatics Technology Co. Ltd (Beijing, China).

2.3. Assembly, annotation, and submission of plastid genomes

Original fluorescence images of the 28 individuals, successfully obtained from the Illumina platform, were transformed to raw data using base calling and these short reads were recorded in the FASTQ format (Chen et al., 2018). Subsequently, low-quality regions in the original sequencing data were filtered using Trimmomatic v.0.32 with default settings (Bolger et al., 2014). Paired-end reads from the clean data were assembled into contigs using the GetOrganelle toolkit (Jin et al., 2020). Finally, 28 plastid genomes were successfully assembled and annotated for the nine *Rosocea* species and one variety (Table 1). For the new plastid genomes, *de novo* assembly graphs were visualized and edited using Bandage, and a whole or nearly whole circular plastid genome was generated (Wick et al., 2015). Using the plastid genome of *R. tibetica* (NC_047420) downloaded from NCBI (National Centre for Biotechnology Information, https://www.ncbi.nlm.nih.gov/) as a reference, the new sequences that had been successfully assembled were annotated using Geneious 11.0.2 with manual correction (Kearse et al., 2012). All the plastid genomes for *Roscoea* species, newly obtained in the present study, were submitted to the NCBI after being checked and found without error. Circular genome visualization was performed using the online tool OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html). In addition, the ITSx program was used to assemble ITS (ITS1, 5.8S rDNA, and ITS2) sequences in the Linux system, and their boundaries were defined based on *R. tibetica* (KM384818) in the NCBI database using Geneious 11.0.2 with manual correction for subsequent species identification analysis (Bengtsson-Palme et al., 2013).

2.4. Analysis on SSR, and expansion and contraction of IR regions

The distribution of simple sequence repeats (SSR) in the new genomes was explored using the search tool MISA (Thiel et al., 2003). The minimum thresholds were set to 10 repeat units for mononucleotide SSRs; five repeat units for dinucleotide; four repeat units for trinucleotide; and three repeat units for tetranucleotide, pentanucleotide, and hexanucleotide SSRs (Murat et al., 2011). Furthermore, the data were uploaded to the IRscope online site (https://irscope.shinyapps.io/irapp/) to analyze the expansion and contraction of the IR/SC boundaries with minor adjustments (Amiryousefi et al., 2018).

2.5. Comparative analysis of the plastid genomes

In this study, a comparative plot consisting of full alignments of the plastid genomes with annotations was produced using mVISTA (https://genome.lbl.gov/vista/mvista), which used the Shuffle-LAGAN model with *Roscoea tibetica* (NC_047420) as the reference. Subsequently, multiple sequence alignments of the plastid genomes from *Roscoea* species were performed using MAFFT v.7.129 at default settings (Katoh and Standley, 2013). DNASP v.6.11 was adopted to compare the sequence divergence among the 28 plastid genomes (Rozas et al., 2017). The step size was set to 200 bp with a 600 bp window length. In addition, DNASP software was used to identify and quantify insertion/deletions (indels),

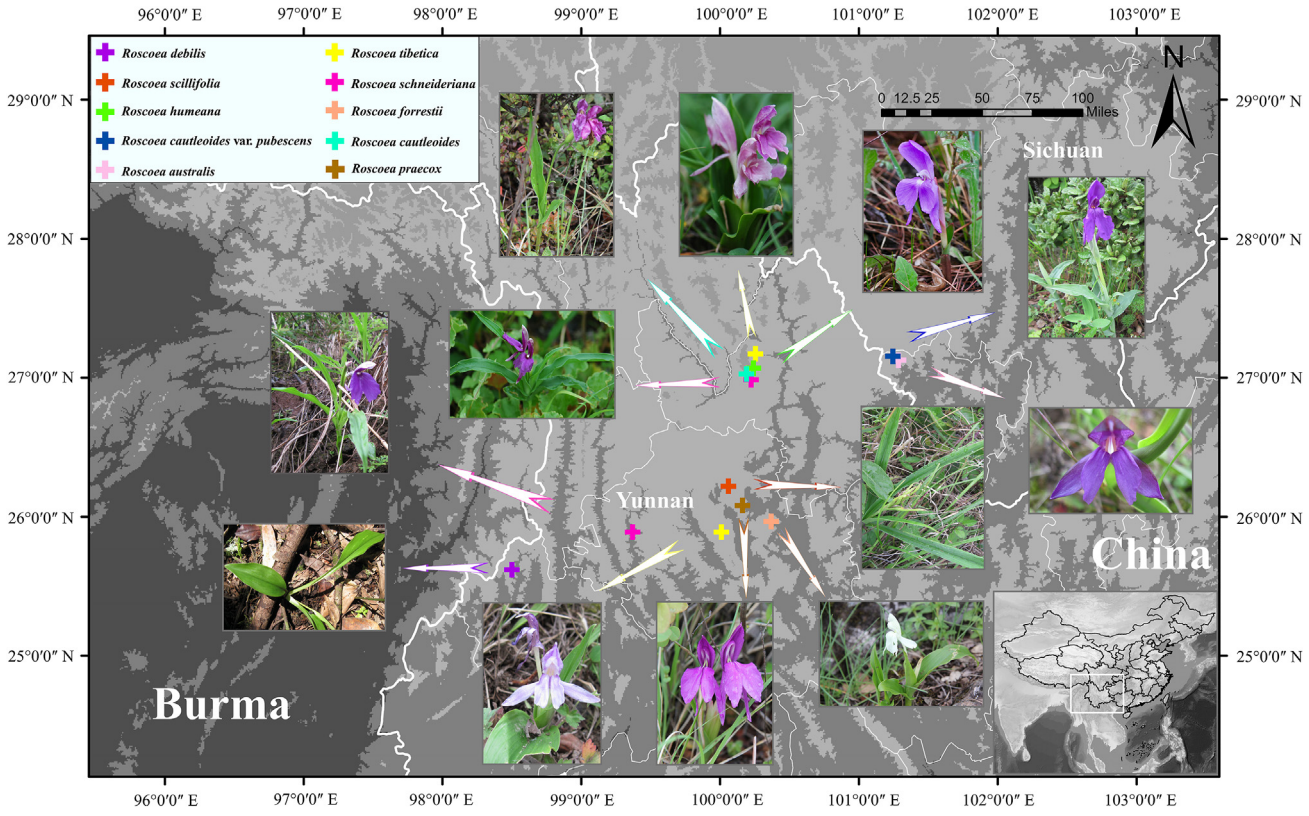


Fig. 1. Sampling distributions of nine species and one variety belonging to the "Chinese" clade of *Roscoea*.

Table 1
Basic characteristics of the 28 plastid genomes of the nine species and one variety belonging to the "Chinese" clade of *Roscoea*.

Species	Total length (bp)	Large single copy (bp)	Small single copy (bp)	Inverted repeat (IR, bp)	GC%	Number of genes	Accession number
Roscoea debilis	163,565	87,945	16,048	29,786	36.0	113	MZ569043
	163,632	88,006	16,054	29,786	36.0	113	MZ569044
R. scillifolia	163,427	87,796	16,071	29,780	36.0	113	MZ561530
	163,426	87,793	16,073	29,780	36.0	113	MZ561529
R. humeana	163,064	87,671	15,949	29,722	36.1	113	OP219811
	163,084	87,692	15,948	29,722	36.1	113	MZ569048
	163,063	87,671	15,948	29,722	36.1	113	MZ569047
R. cautleoides var. *pubescens*	163,783	88,074	16,135	29,787	36.0	113	OP219812
	163,796	88,135	16,075	29,793	36.0	113	MZ569040
	163,795	88,134	16,075	29,793	36.0	113	MZ569041
	163,785	88,075	16,136	29,787	36.0	113	MZ569039
R. tibetica	163,529	88,042	15,861	29,813	36.1	113	OP219813
	163,529	88,042	15,861	29,813	36.1	113	MZ618239
	163,530	88,043	15,861	29,813	36.1	113	MZ561531
R. schneideriana	163,423	87,772	16,057	29,797	36.1	113	OP219814
	163,423	87,772	16,057	29,797	36.1	113	OP219815
	163,483	87,818	16,113	29,776	36.0	113	MZ569051
	163,389	87,776	16,041	29,786	36.1	113	MZ569050
R. forrestii	163,616	88,139	15,843	29,817	36.1	113	MZ569045
	163,642	88,157	15,849	29,818	36.1	113	MZ569046
R. cautleoides	163,089	87,702	15,943	29,722	36.1	113	MZ569042
	163,089	87,702	15,943	29,722	36.1	113	MZ569038
R. praecox	163,419	87,779	16,052	29,794	36.0	113	MZ569049
	163,420	87,780	16,052	29,794	36.0	113	OP219810
	163,420	87,759	16,051	29,805	36.0	113	OP219809
	163,420	87,759	16,051	29,805	36.0	113	MZ561528
R. australis	163,590	87,978	16,050	29,781	36.0	113	MZ569037
	163,579	87,966	16,051	29,781	36.0	113	MZ569036

mutations, parsimony information, and nucleotide variability (Pi) in all aligned datasets.

2.6. Phylogenetic analysis based on plastid genomes

The 28 plastid genomes of *Roscoea* were used to construct phylogenetic trees based on Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI). These plastid genome sequences were aligned using MAFFT v.7.129 at default settings (Katoh and Standley, 2013). The alignment was conducted with the MP method using MEGA v.7.0.26, with 1000 bootstrap replicates (Kumar et al., 2016). For ML and BI, jModelTest v.2.1.7 was used to test the best substitution model for the Akaike information criterion (AIC), and then GTR + I + G was chosen (Darriba et al., 2012). The ML analysis was performed using RAxML v.8.2.4 (Stamatakis, 2014), and 1000 replications were set to calculate the bootstrap probability of each branch. The BI was conducted in MrBayes v.3.2.6 (Ronquist et al., 2012), and the Markov Chain Monte Carlo (MCMC) algorithm was calculated for 1,000,000 generations with a sampling tree every 1000 generations. The first 25% of generations were discarded as burn-in. The state was considered to be reached when the average standard deviation of the split frequencies was < 0.01, and a consensus tree was constructed using the remaining trees.

2.7. Species discrimination using plastid genome

To identify species based on the plastid genomes (DNA superbarcode) of *Roscoea* species, three basic methods, namely Blast, Distance, and Tree-building, were adopted to compare the success rates of plastid genomes, intergenic spacer regions (IGS), coding sequences (CDS), hypervariable regions (HVR), and universal DNA barcodes (ITS, *matK*, *rbcL*, and *trnH-psbA*). For the Blast method, sequences corresponding to all individuals in the five types of datasets were used as query sequences with $E < 1 \times 10^{-5}$ to construct species databases of corresponding types, and Blast software was applied to compare corresponding sequences of each individual in the constructed database to determine whether the sequence with the highest similarity came from the same species. For the Distance method, MEGA v.7.0.26 was used to calculate the genetic distances between the species in *Roscoea* (Kumar et al., 2016). When the minimum genetic distance between one species and any other species was larger than the maximum distance among individuals within the species, the species was considered to be successfully identified. For the Tree-building method, a neighbor-joining tree (NJ) based on the five datasets was constructed using MEGA v.7.0.26, with two species of *Curcuma* L. (*C. longa*: MK965541; *C. phaeocaulis*: MK621772) for the plastid genome, and two species in *Zingiber* Boehm. (*Z. wrayi*: HM236155; *Z. kerrii*: MT793687) used for the ITS downloaded from NCBI as the outgroup.

3. Results

3.1. Organization of the plastid genomes of species in the "Chinese" clade of Roscoea

In total, 28 plastid genome sequences were newly obtained from nine species and one variety in the "Chinese" clade of *Roscoea*, and these genomes shared similar structures and organization. The genome sizes of these species ranged from 163,063 bp to 163,796 bp, comprising a typical quadripartite structure with two inverted repeat regions (IRa and IRb, 29,722–29,818 bp) separated by a large single copy (LSC, 87,671–88,157 bp) and a small single copy (SSC, 15,843–16,139 bp) (Fig. S1, Table 1). All plastid genomes had 113 genes, including 79 protein-coding genes, 30 tRNA genes, and four rRNA genes. Here, we did not count the repeated genes in the IR regions. Among all the genes, there were 18 intron-containing genes, of which 14 genes contained only one intron (*trnK-UUU*, *rps16*, *trnG-GCC*, *atpF*, *rpoC1*, *trnV-UAC*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnL-GAU*, *trnA-UGC*, and *ndhA*) and four genes contained two introns (*ycf3*, *trnL-UAA*, *clpP*, and *rps12*) (Table 2). The GC contents among the genomes were also similar, ranging from 36.0% to 36.1%. Moreover, numerous SSR loci were found in these genomes through MISA analysis. Most of the mononucleotide SSRs were composed of A/T mofits and the dinucleotide ones were composed of AT/TA. SSR was mainly located in the LSC region (82–106), followed by SSC (23–31), and then IR (12–21) (Fig. 2).

3.2. Expansion and contraction of the IR region

In the present study, four boundaries of the plastid genomes, namely JLB (LSC-IRb), JSB (SSC–IRb), JSA (SSC-IRa), and JLA (LSC-IRa) were relatively conserved among the species in *Roscoea*. The IR/SC junctions of the 28 plastid genomes mainly contained five

Table 2
A list of genes found in the plastid genomes of *Roscoea* species.

Category for gene	Group of genes	Name of genes
Self-replication	Large subunit of ribosome	*rpl2*[ab], *rpl14*, *rpl16*[b], *rpl20*, *rpl22*, *rpl23*[a], *rpl32*, *rpl33*, *rpl36*
	Small subunit of ribosome	*rps2*, *rps3*, *rps4*, *rps7*[a], *rps8*, *rps11*, *rps12*[ab], *rps14*, *rps15*, *rps16*[b], *rps18*, *rps19*[a]
	DNA dependent RNA polymerase	*rpoA*, *rpoB*, *rpoC1*[b], *rpoC2*
	rRNA gene	*rrn4.5*[a], *rrn5*[a], *rrn16*[a], *rrn23*[a]
	tRNA gene	*trnI-CAU*[a], *trnL-CAA*[a], *trnV-GAC*[a], *trnI-GAU*[ab], *trnA-UGC*[ab], *trnR-ACG*[a], *trnN-GUU*[a], *trnL-UAG*, *trnP-UGG*, *trnW-CCA*, *trnM-CAU*, *trnV-UAC*[b], *trnF-GAA*, *trnL-UAA*[b], *trnT-UGU*, *trnS-GGA*, *trnfM-CAU*, *trnG-GCC*[b], *trnS-UGA*, *trnT-GGU*, *trnE-UUC*, *trnY-GUA*, *trnD-GUC*, *trnC-GCA*, *trnR-UCU*, *trnG-UCC*, *trnS-GCU*, *trnQ-UUG*, *trnK-UUU*[b], *trnH-GUG*[a]
Gene for photosynthesis	Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Subunits of NADH-dehydrogenase	*ndhA*, *ndhB*[ab], *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Subunits of cytochrome b/f complex	*petA*, *petB*[b], *petD*[b], *petG*, *petL*, *petN*
	Subunit for ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[b], *atpH*, *atpI*
	Large subunit of rubisco	*rbcL*
Other genes	Translational initiation factor	*infA*
	Maturase	*matK*
	Protease	*clpP*[b]
	Envelope membrane protein	*cemA*
	Subunit of Acetyl-carboxylase	*accD*
	C-type cytochrome synthesis gene	*ccsA*
	Open reading frames (ORF, ycf)	*ycf1*[a], *ycf2*[a], *ycf3*[b], *ycf4*

Note: The "a" label after gene names reflects genes located in IR regions. Intron containing gene is indicated by "b".

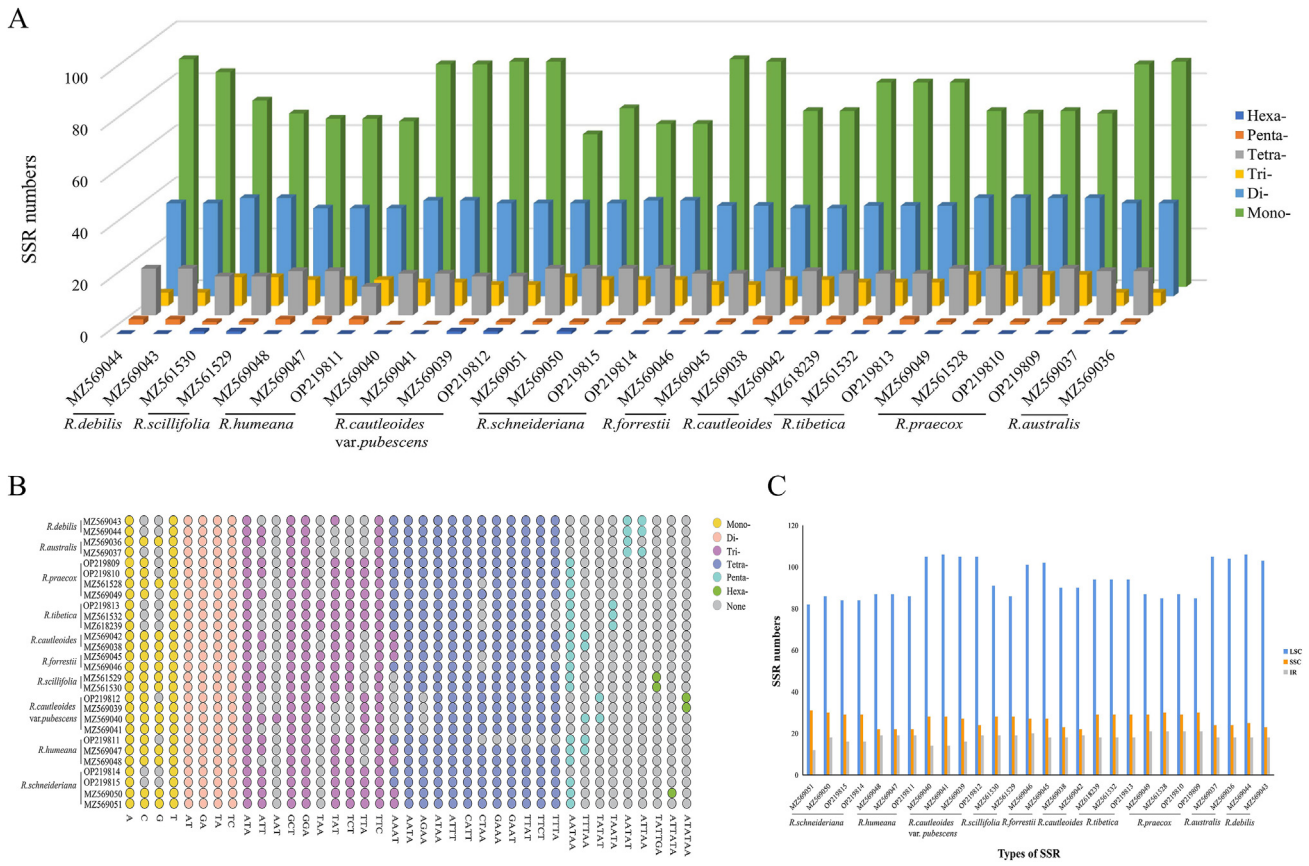


Fig. 2. Analysis of simple sequence repeats (SSRs) in the 28 plastid genomes. (A) Number of different SSR types detected in the genomes; (B) SSR motifs in different repeat types; (C) number of SSRs identified in LSC, SSC, and IR regions.

genes (*rpl22*, *rps19*, *ycf1*, *ndhF*, and *psbA*) (Fig. 3). The *rpl22* gene was completely present within the LSC, at a distance of 25–88 bp to the JLB junction. In both *R. schnideriana* and *R. praecox*, *rps19* distance to the JLA junction varied. In other species, we only detected interspecies variation in the distance (120–133 bp) to the JLA junction. The *ycf1* gene crossed the JSA junction, expanding 1545–1603 bp into SSC, but a truncated *ycf1* remained in IRb. Furthermore, *psbA* was completely present in the LSC region and farther from IRa (100–136 bp). In total, most of the genes in the four junctions showed contraction, except *ycf1* in the JSA, and the genetic variation possessed adequate intraspecific conservation but significant interspecific differences in *Roscoea* (Fig. 3).

3.3. Hypervariable regions of the plastid genomes

In this study, 28 complete plastid genomes from the "Chinese" clade of *Roscoea* were compared using mVISTA with the *R. tibetica* genome as reference. Comparative analysis showed that the plastid genomes were evolutionarily conserved with similar structures and gene orders. The noncoding regions showed much richer genetic divergence than the coding regions, except the gene *ycf1* (Fig. 4). Meanwhile, sliding window analysis was performed on these plastomes. The results suggested that the nucleotide variability (Pi) of the IR regions was significantly lower than that of the LSC and SSC regions, indicating that the IR regions were more conserved within the whole plastid genome (Fig. 5). Nucleotide diversity can be used to estimate the divergence level of the different regions. Coupled with the computation of the number of variable positions per unit length of the gene, the total variation was 0.00249, while the Pi value ranged from 0 to 0.02764. The sliding window revealed a series of hypervariable regions (HVR), including *atpH-atpI*, *trnS-GCU*, *rpl32-trnL-UAG*, *ycf1*, *petA-psbJ*, *psbB-psbT*, *psaC-ndhE*, *ndhF*, and *trnS-GCU-trnG-GCC*. It is noteworthy to mention that all HVRs were almost located in the LSC and SSC, except the *ycf1* gene; at the same time, these HVRs generally belonged to intergenic regions (IGS) (Fig. 5). Compared to these HVRs, the plastid genome had many more informative sites. There were 1264 mutation sites in the matrix, of which 1212 were parsimony-informative sites (Table 3).

3.4. Plastid phylogenomics of the "Chinese" clade of Roscoea

In the present study, plastid phylogenomics for the species in the "Chinese" clade of *Roscoea* was performed based on MP, ML, and BI analyses. The structures of the phylogenetic trees between the three methods were consistent (Fig. 6). The nine species and one variety were clearly monophyletic, except for *R. debilis*. In this analysis, the nine species and one variety were mainly clustered into four subclades (I, II, III, and IV). Subclade I, consisting of *R. tibetica* and *R. forrestii*, was the first to be separated from the other *Roscoea* species. Subclade II was composed of three species that included only the paraphyletic group, namely *R. debilis*. Finally, subclades III and IV were sister groups, composed of three and two species, respectively. It should be noted that *R. cautleoides* and its variety, *R. cautleoides* var. *pubescens*, was clustered into different subclades based on all the methods.

3.5. Species discrimination of plastid genome and other regions

We evaluated the success rate at which various DNA barcodes identify species in the "Chinese" clade of *Roscoea*. Specifically, we

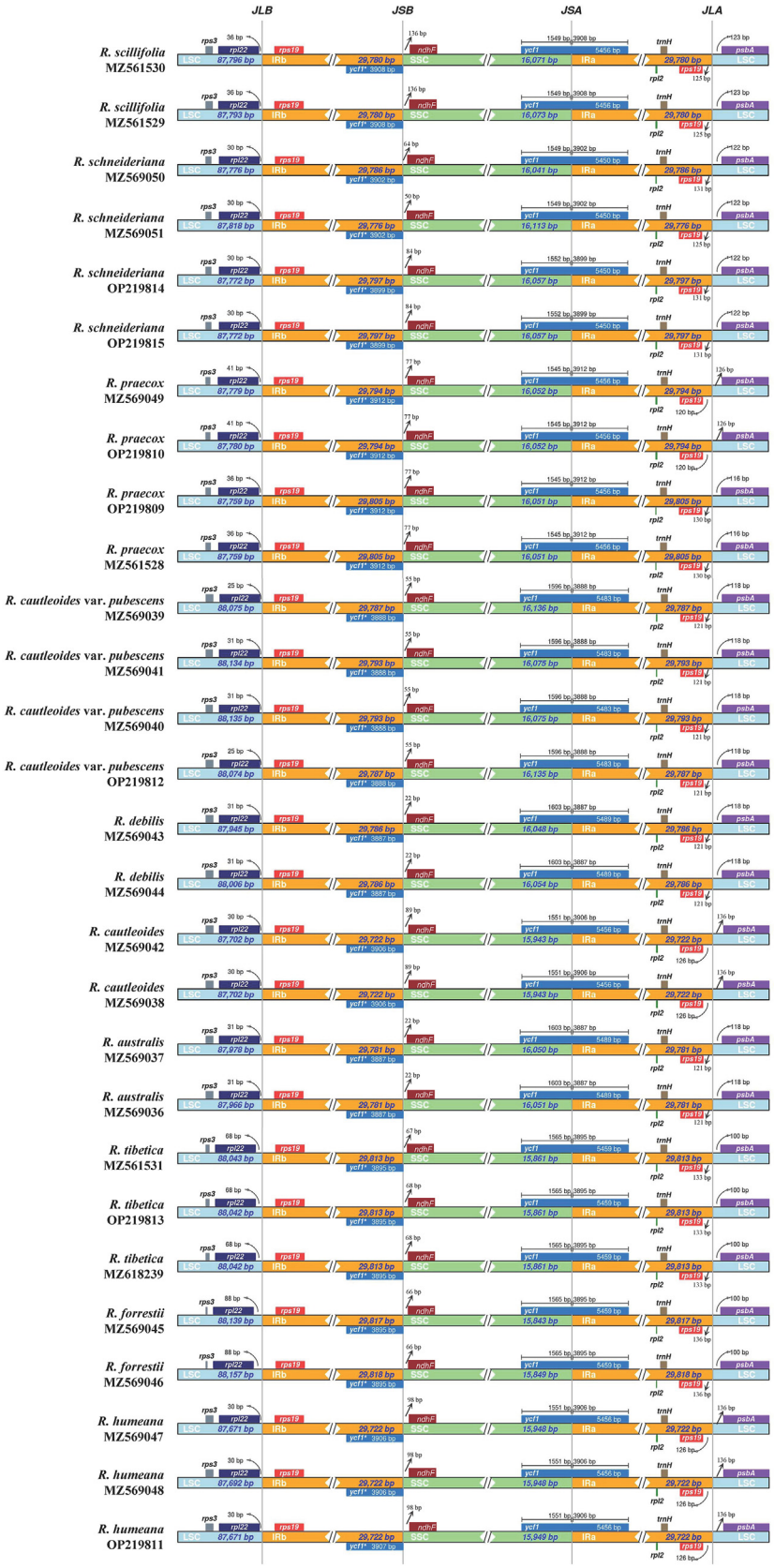


Fig. 3. Comparison of junctions of LSC, SSC, and IR regions in the plastid genomes among the nine species and one variety in *Roscoea*. The "*ycf1**" gene with an asterisk mark refers to the truncated *ycf1* in IRb.

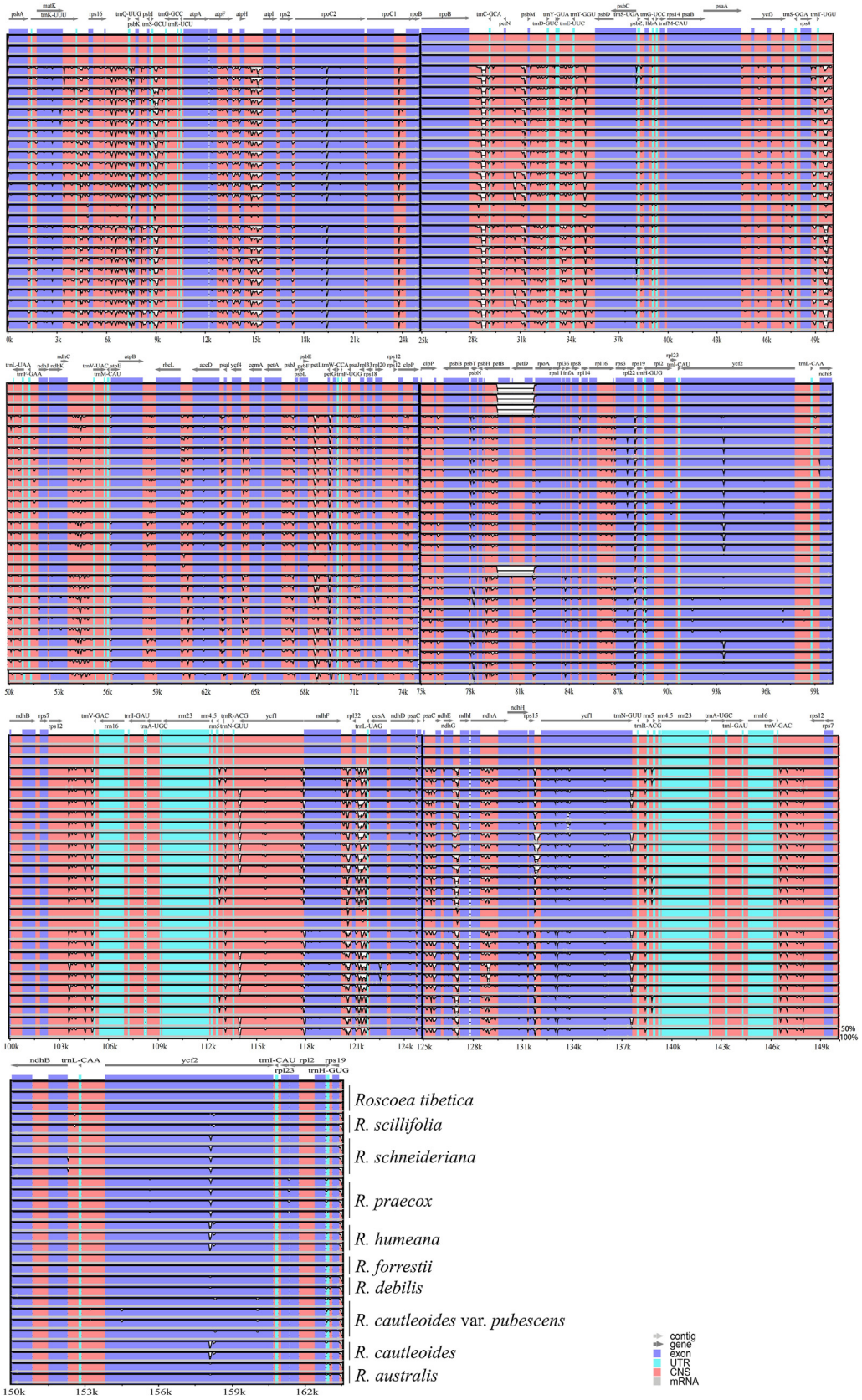


Fig. 4. Sequence alignment of the 28 plastid genomes performed using the mVISTA program with *Roscoea tibetica* (NC_047420) as a reference. The gray arrow and its appearance represent the direction and position of gene, respectively. The y-axis indicates a percent identity between 50% and 100%, and the red and blue areas indicate intergenic and gene regions, respectively.

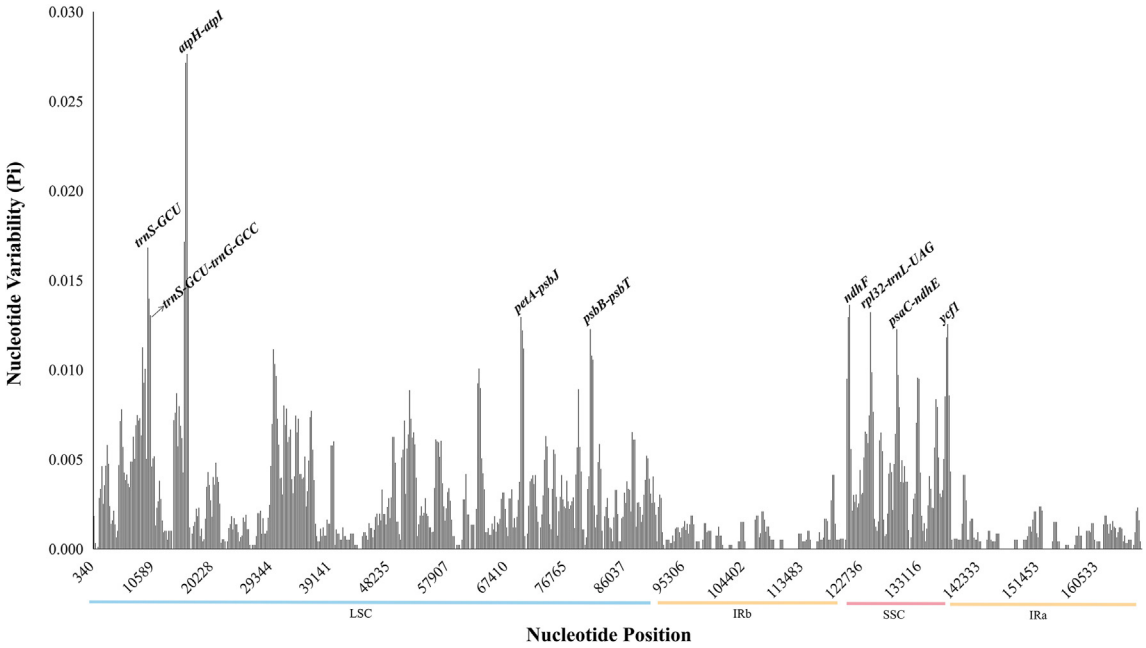


Fig. 5. Sliding window analysis of the plastid genomes of *Roscoea* species.

Table 3
Comparison of variable sites among plastid genomes, IGS, CDS, each HVR, and universal DNA barcodes in *Roscoea* species.

	No. sites	No. variable sites	No. parsimony information sites	No. mutations	No. InDels	Nucleotide diversity (Pi)
Genome	167,805	1255	1212	1264	4096	0.00249
IGS	53,939	728	710	736	3758	0.00500
CDS	82,554	369	379	370	367	0.00141
atpH-atpI	1370	40	39	41	141	0.01484
trnS-GCU	88	1	1	1	0	0.00346
ndhF	2218	21	20	21	11	0.00289
petA-psbJ	900	39	39	40	70	0.01402
psaC-ndhE	844	24	24	25	146	0.01246
psbB-psbT	204	6	5	6	28	0.00918
rpl32-trnL-UAG	1116	40	40	40	92	0.02138
trnS-GCU-trnG-GCC	976	23	23	24	172	0.01089
ycf1	5502	45	43	45	60	0.00279
ITS	568	33	25	33	3	0.01445
matK	1524	18	13	19	24	0.00305
rbcL	1462	2	2	2	0	0.00055
trnH-psbA	657	6	6	6	43	0.00268

used Blast, Distance, and Tree-building approaches with five data sets as DNA barcodes: whole plastid genomes, intergenic spacer regions, coding sequences, highly variable regions, and universal DNA barcodes. Among these methods, Tree-building based on NJ tended to provide a higher success rate of species identification than the others for the genome, CDS, and IGS datasets; nevertheless, Blast showed the best performance for all the HVR regions (Fig. 7). The dataset of the plastid genomes, as a DNA super-barcode, had an aligned length of 169,263 sites when the outgroups were included. The plastid genomes showed the highest power of species identification in *Roscoea* (90%) with strong support; in other words, the plastid genomes could successfully identify all the species, except *R. debilis* (Fig. S2). IGS also possessed a similar discriminatory power as the genomes (90%), followed by CDS (80%). Among the hypervariable regions and universal DNA barcodes, *ycf1* and ITS had the highest success rate (70% and 50%) according to the Tree-building method. The species discrimination ability for each of these were *atpH-atpI, petA-psbJ*, *psaC-ndhE*, *matK*, and *rbcL* (40%); *rpl32-trnL-UAG* and *trnS-GCU-trnG-GCC* (30%); *ndhF* (20%), and *trnS-GCU* and *trnH-psbA* (0%) (Figs. 7 and S3).

4. Discussion

4.1. Conservatism and divergence of the plastid genomes among species in the "Chinese" clade of Roscoea

Generally, the size of the plastid genomes in the photosynthetic land plants ranges from 120 kb to 160 kb, and comprises 100–120 unique genes (Wiche et al., 2011). According to the present study, the species in the "Chinese" clade of *Roscoea* possess sizes of plastid genomes from 163,063 bp in *R. humeana* to 163,796 bp in *R. cautleoides* var. *pubescens*, which are clearly longer than those of *Veratrum* in Melanthiaceae (Zhang et al., 2022b), *Bulbophyllum* in Orchidaceae (Tang et al., 2021), and *Fritillaria* in Liliaceae (Chen et al., 2022), but similar to *Curcuma* in Zingiberaceae (Liang and Chen, 2021). This indicates that the plastid genomes in Zingiberaceae within monocots are long. Plastome structure and gene content are highly conserved in land plants, except for heterotrophic plants (Bai et al., 2021; Wu et al., 2020). However, *inf*A, the most common gene lost in angiosperm plastid genomes, is revealed in *Roscoea* species which has also been reported to occur in the

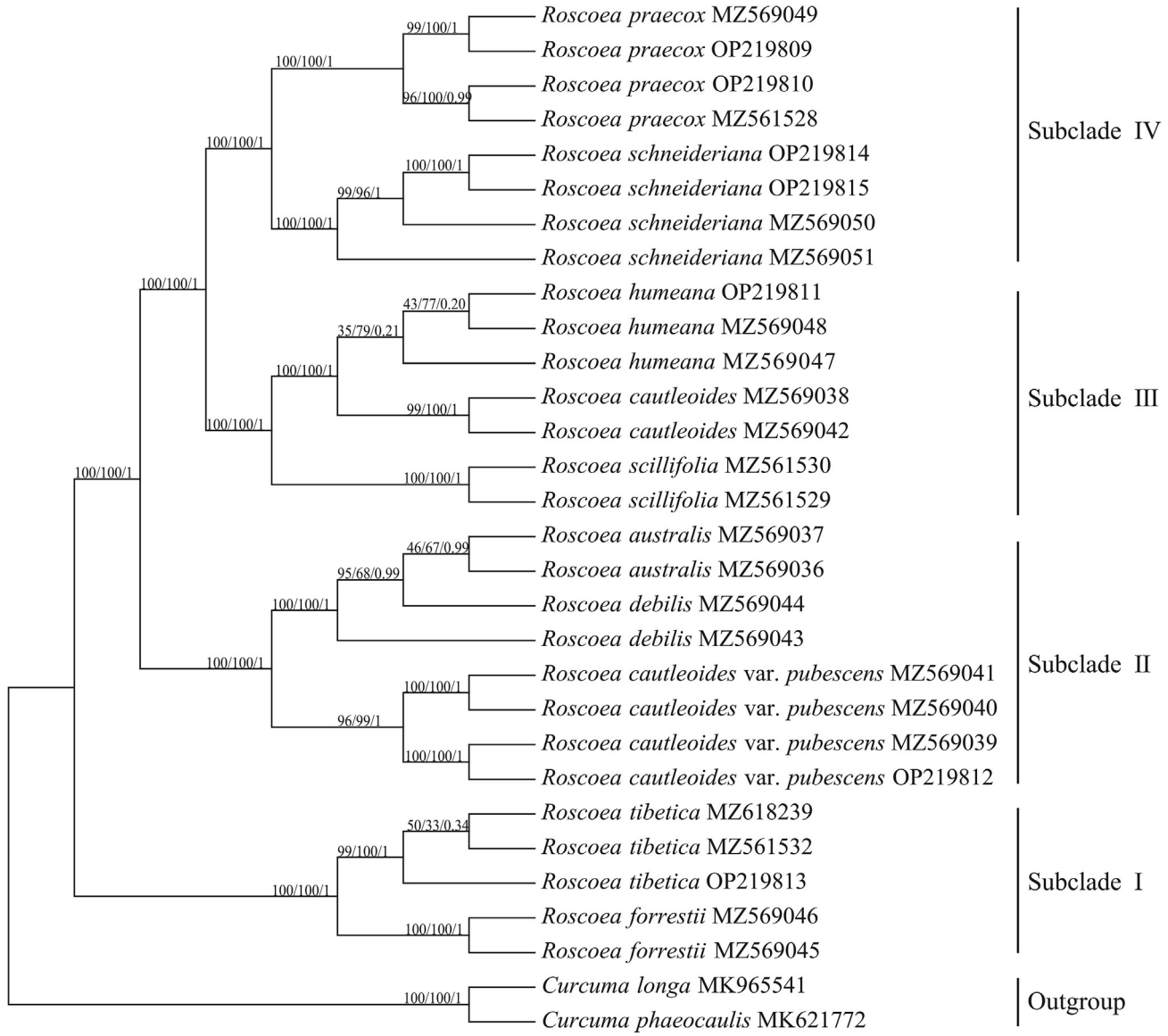


Fig. 6. Plastid phylogenomics of the "Chinese" clade of *Roscoea* using Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) methods based on the plastid genome of *Roscoea* species. MP and ML bootstrap support values (BS)/BI posterior probability values (PP) are respectively shown at the nodes.

majority of species in *Veratrum* L. (Zhang et al., 2022b). In this study, 113 genes were annotated across all the genomes of the nine species and one variety in *Roscoea*, including 79 protein-coding genes, 30 tRNA genes, and four rRNA genes that showed good conservatism across species in this genus.

SSR markers from plastid genomes are useful tools for evaluating genetic diversity and revealing phylogeography because of their high polymorphism and good reproducibility (Mohammad-Panah et al., 2017; Yang et al., 2021; Zhang et al., 2022b). Herein, the numbers of mononucleotide and dinucleotide repeats were 59–88 and 28–38, respectively, representing most of the SSRs in all the *Roscoea* species (Fig. 2A). Among these repetitive sequences, A/T and AT/TA repeat units accounted for approximately 90% of all SSR sites (Fig. 2B), corresponding to the plastid genome composition of angiosperms (Tang et al., 2021; Zhang et al., 2022b). Meanwhile, most of the SSR were located in the LSC and SSC (Fig. 2C), revealing conservatism of the IR regions, which was also supported by the distributions of the hypervariable regions (HVR) (Fig. 5). Concurrently, analysis of the SSR loci in the plastid genomes of *Roscoea* is also helpful for designing SSR primers, thereby providing an effective means for studies on topics such as population genetic diversity, phylogeography, speciation, and species identification etc (Powell et al., 1995; Li et al., 2020b; Tang et al., 2021).

Expansion and contraction at the boundaries of the IR regions of the plastid genome are important factors that cause size variations and play a major role in structural stability and evolution (Shahzadi et al., 2019). In this study, the gene distributions at the IRb/SSC and SSC/IRa boundaries in plants were quite similar, and each had one *ycf1* gene, with full-length *ycf1* at the SSC/IRa boundary of 5450–5489 bp. In contrast, the IRb/SSC boundary contained a truncated *ycf1* gene of only 3887–3912 bp. In addition, the boundary of *Roscoea* was expanded, *rps19* and *trnH* were located in the IRa region, and both genes changed from one to two copies. This is consistent with the boundary genes of *Curcuma*, *Amomum*, and *Zingiber* in Zingiberaceae (Liang et al., 2020). The results indicated that the boundary between the IR region and the LSC and SSC regions is highly conserved. Overall, the expansion and contraction of the IR boundaries in this study indicated obvious intraspecific

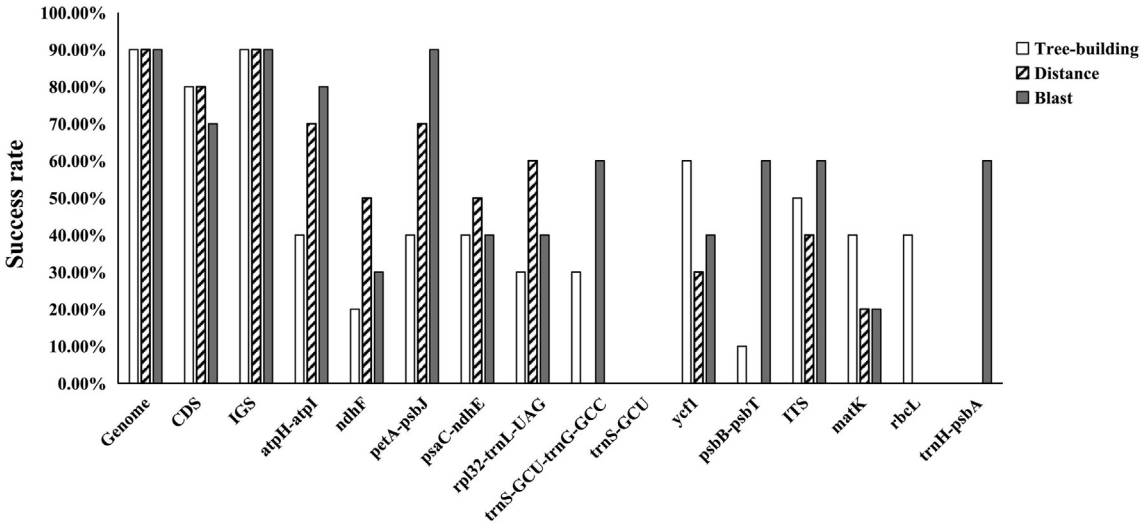


Fig. 7. Success rates of five data sets at identifying species belonging to the "Chinese" clade in *Roscoea*. Barcoding sequences include the whole plastid genome, plastid coding sequences (CDS), intergenic spacer regions (IGS), highly variable regions (HVRs), and universal DNA barcodes. Three methods were used to analyze these sequences: Blast, Distance, and Tree-building.

conservation and interspecific variability that afforded reliable species discrimination in *Roscoea* species.

4.2. Plastid phylogenomics of the "Chinese" clade of Roscoea

Roscoea has been adequately shown to consist of two independent clades with significant biogeographic disjunction based on molecular (Ngamriababsakul et al., 2000; Zhao et al., 2017) and ecological pieces of evidences (Li et al., 2020a). Although Zhao et al. (2017) constructed a preliminary species-level phylogeny based on ITS sequences and plastid DNA regions, phylogenetic relationships within the genus, especially for the "Chinese" clade (namely the NIC clade in the article), were still poorly resolved because of limited sampling, insufficient genetic sites, hybridization, incomplete lineage sorting, and budding speciation. In the present study, we provided a more informative phylogeny of the group that was composed of four distinct subclades (Fig. 6).

Among these species, *Roscoea tibetica*, a species that occurs on both sides of the Brahmaputra gap (Ngamribabsakul et al., 2000), was first separated from the other species, together with *R. forrestii* (the Subclade I). The species might be the most complicated species in *Roscoea* with extremely wide distributions and complicated morphological variations (Cowley, 2007); therefore, more representative sampling for this species is necessary to further verify its phylogenetic position. It is interesting that the two individuals of *R. debilis* were not gathered into an independent group, but instead formed a paraphyletic group with *R. australis*, although they were easily discriminated according to morphological characteristics (Fig. 1). In addition, *R. cautleoides* var. *pubescens* was distinctly separated from the original variety of *R. cautleoides*, further supporting our previous conclusion that recovered the variety into an independent species, namely *R. pubescens* (Zhang et al., 2015). Herein, *R. humeana* and *R. cautleoides* were sister species in phylogeny, which was consistent with a previous study (Zhao et al., 2017) and was also demonstrated to have diverged recently, coincident with the Quaternary climate cycle (Zhao et al., 2016b). The close relationship between the two species was also verified by the interspecific hybridization that occurred in sympatric distributions (Du et al., 2012). Indeed, the present plastid phylogenomics could not thoroughly resolve the phylogenetic relationships among species in the "Chinese" clade of *Roscoea* probably due to the controversial species definitions, insufficient representative sampling, and complicated speciation processes of these species (Zhao et al., 2017). Nevertheless, plastid phylogenomics has advanced our understanding of the phylogeny of *Roscoea*.

4.3. Species discrimination on species of the "Chinese" clade of Roscoea using plastid genome

Roscoea species are difficult to identify due to complicated morphological variations, interspecific hybridization, and budding speciation (Du et al., 2012; Zhao et al., 2017, 2021). Here, we used plastid genomes as a super barcode to identify nine species and one variety in the "Chinese" clade of *Roscoea*. The super barcode identified all species in *Roscoea* except *R. debilis* (90%), showing similar success rate to universal barcodes (ITS + *trnH-psbA*), but with much better reliability (Zhang et al., 2014). This finding is consistent with previous studies that have shown plastid genomes can significantly improve the reliability of species identification compared to universal barcodes (Ji et al., 2019; Chen et al., 2020, 2022). For example, super barcodes have also successfully been used to discriminate most *Fritillaria* species in China (20/21), except for *F. cirrhosa*, which has a complicated lineage (Chen et al., 2022). In *Panax*, *Berberis*, and *Rhododendron*, plastid genomes have been shown to increase the discriminatory power of barcodes, but they were still not powerful enough to accurately identify all species (Ji et al., 2019; Kreuzer et al., 2019; Fu et al., 2022). The discriminatory power of plastid genomes is likely limited by the maternal inheritance of plastids (Nock et al., 2015; Li et al., 2015). Thus, future DNA barcoding efforts should include biparentally inherited nuclear genes.

Note that intergenic spacer regions also identified *Roscoea* species at a high rate (90%), probably because these non-coding regions mostly consist of highly variable sequences, thereby providing more informative characters (Small et al., 1998; Zheng et al., 2013; Yang et al., 2020). Although the plastid genome contains many noncoding regions, relatively few have been exploited for studies on interspecific phylogeny and intraspecific phylogeography (Shaw et al., 2007). Taken together, our findings indicate that super barcodes are superior to universal barcodes for species identification of complicated or recently diverged species, but may not reliably identify all problematic species.

5. Conclusion

In the present study, we first analyzed the basic characteristics of the plastid genomes of species belonging to the "Chinese" clade of *Roscoea*, and revealed good conservatism of genome size, gene content, SSR repeats, and IR boundaries among the species. Herein, the expansion and contraction of the IR regions showed obvious intraspecific conservatism and interspecific variability in the nine species and one variety, which provided efficient tools for exploring phylogenetic relationships and discriminating the species in *Roscoea*. Plastid phylogenomics revealed well-supported phylogenetic relationships among species in the "Chinese" clade of *Roscoea*. Meanwhile, the plastid genomes could afford a much better ability for species identification than universal barcodes (*rbcL*, *matK*, ITS, *trnH-psbA* etc.), especially concerning the reliability in this study. However, plastid genomes still can't completely resolve the complicated phylogenetic relationships and species discrimination for such a special group with complicated speciation history and possessing confused species definition.

Author contributions

DQZ conceived the research, collected the molecular materials and voucher specimens; HSH performed experiments, data analysis, and wrote initial draft of this manuscript; JYM, XW, and YZL took part in amount of works in data analysis and revision of the manuscript; BJ participated in field works for some species and gave revised suggestions for the manuscript; and then, DQZ revised this manuscript and submitted it finally. All authors revise the manuscript and approve the final version.

Declaration of competing interest

The author declares no conflict of interest.

Acknowledgements

This study was supported by National Natural Science Foundation of China (32060091 & 31660081), Reserve Talents Project for Young and Middle-Aged Academic and Technical Leaders of Yunnan Province (202105AC160063).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pld.2023.03.012.

References

Amiryousefi, A., Hyvönen, J., Poczai, P., 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34, 3030–3031.
Bai, H.R., Oyebanji, O., Zhang, R., et al., 2021. Plastid phylogenomic insights into the evolution of subfamily Dialioideae (Leguminosae). Plant Divers. 43, 27–34.
Bengtsson-Palme, J., Ryberg, M., Hartmann, M., et al., 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol. Evol. 4, 914–919.
Bolger, A.M., Marc, L., Bjoern, U., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Chen, Q., Wu, X.B., Zhang, D.Q., 2020. Comparison of the abilities of universal, super, and specific DNA barcodes to discriminate among the original species of Fritillariae cirrhosae bulbus and its adulterants. PLoS One 15, e0229181.
Chen, Q., Hu, H.S., Zhang, D.Q., 2022. DNA barcoding and phylogenomic analysis of the genus *Fritillaria* in China based on complete chloroplast genomes. Front. Plant Sci. 13, 764255.
Chen, S.F., Zhou, Y.Q., Chen, Y.R., et al., 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.
Cowley, E.J., 2007. The Genus *Roscoea*. The Royal Botanic Gardens, Kew, UK.
Darriba, D., Taboada, G.L., Doallo, R., et al., 2012. jModelTest 2: more models, new heuristics and high-performance computing. Nat. Methods 9, 772.
Du, G.H., Zhang, Z.Q., Li, Q.J., 2012. Morphological and molecular evidence for natural hybridization in sympatric population of *Roscoea humeana* and *R. cautleoides* (Zingiberaceae). J. Plant Res. 125, 595–603.
Doyle, J.J., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.
Escobari, B., Borsch, T., Quedensley, T.S., et al., 2021. Plastid phylogenomics of the Gynoxoid group (Senecioneae, Asteraceae) highlights the importance of motif-based sequence alignment amid low genetic distances. Am. J. Bot. 108, 2235–2256.
Fu, C.N., Wu, C.S., Ye, L.J., et al., 2019. Prevalence of isomeric plastomes and effectiveness of plastome super-barcodes in yews (*Taxus*) worldwide. Sci. Rep. 9, 2773.
Fu, C.N., Mo, Z.Q., Yang, J.B., et al., 2022. Testing genome skimming for species discrimination in large taxonomically difficult genus *Rhododendron*. Mol. Ecol. Resour. 22, 404–414.
Ji, Y.H., Liu, C.K., Yang, Z.Y., et al., 2019. Testing and using complete plastomes and ribosomal DNA sequences as the next generation DNA barcodes in *Panax* (Araliaceae). Mol. Ecol. Resour. 19, 1333–1345.
Jin, J.J., Yu, W.B., Yang, J.B., et al., 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21, 241.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance. Mol. Biol. Evol. 30, 772–780.
Kearse, M., Moir, R., Wilson, A., et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Kleinwee, I., Larridon, I., Shah, T., et al., 2022. Plastid phylogenomics of the Sansevieria clade of *Dracaena* (Asparagaceae) resolves a recent radiation. Mol. Phylogenet. Evol. 169, 107404.
Kreuzer, M., Howard, C., Adhikari, B., et al., 2019. Phylogenomic approaches to DNA barcoding of herbal medicines: developing clade-specific diagnostic characters for *Berberis*. Front. Plant Sci. 10, 586.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Li, D.B., Ou, X.K., Zhao, J.L., et al., 2020a. An ecological barrier between the Himalayas and the Hengduan Mountains maintains the disjunct distribution of *Roscoea*. J. Biogeogr. 47, 326–341.
Li, S., Liu, S.L., Pei, S.Y., et al., 2020b. Genetic diversity and population structure of *Camellia huana* (Theaceae), a limestone species with narrow geographic range, based on chloroplast DNA sequence and microsatellite markers. Plant Divers. 42, 343–350.
Li, X.W., Yang, Y., Henry, R.J., et al., 2015. Plant DNA barcoding: from gene to genome. Biol. Rev. 90, 157–166.
Li, Z.Z., Gichira, A.W., Muchuku, J.K., et al., 2021. Plastid phylogenomics and biogeography of the genus *Monochoria* (Pontederiaceae). J. Syst. Evol. 59, 1027–1039.
Liang, H., Zhang, Y., Deng, J.B., et al., 2020. The complete chloroplast genome sequences of 14 *Curcuma* species: insights into genome evolution and phylogenetic relationships within Zingiberales. Front. Genet. 11, 802.
Liang, H., Chen, J., 2021. Comparison and phylogenetic analyses of nine complete chloroplast genomes of Zingibereae. Forests 12, 710.
Luo, M.H., Wan, H.L., Lin, H.H., 2008. Species and distribution of *Roscoea* in China and the medicinal use. Chin. Wild Plant Resour. 27, 35–37+41.
Misra, A., Srivastava, S., Verma, S., et al., 2015. Nutritional evaluation, antioxidant studies and quantification of poly phenolics in *Roscoea purpurea* tubers. BMC Res. Notes 8, 324.
Mohammad-Panah, N., Shabanian, N., Khadivi, A., et al., 2017. Genetic structure of gall oak (*Quercus infectoria*) characterized by nuclear and chloroplast SSR markers. Tree Genet. Genomes 13, 70.
Murat, C., Riccioni, C., Belfiori, B., et al., 2011. Distribution and localization of microsatellites in the Perigord black truffle genome and identification of new molecular markers. Fungal Genet. Biol. 48, 592–601.
Ngamriabsakul, C., Newman, M.F., Cronk, Q.C.B., et al., 2000. Phylogeny and disjunction in *Roscoea* (Zingiberaceae). Edinb. J. Bot. 57, 39–61.
Nock, C.J., Waters, D., Edwards, M.A., et al., 2015. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol. J. 9, 328–333.
Powell, W., Morgante, M., McDevitt, R., et al., 1995. Polymorphic simple sequence repeat regions in chloroplast genomes, application to the population genetics of pines. Proc. Natl. Acad. Sci. U.S.A. 92, 7759–7763.
Rawat, S., Jugran, A.K., Bhatt, I.D., et al., 2018. Influence of the growth phenophases on the phenolic composition and anti-oxidant properties of *Roscoea procera* Wall. in western Himalaya. J. Food Sci. Technol. 55, 578–585.
Ronquist, F., Teslenko, M., Van, D.M.P., et al., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., et al., 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. Mol. Biol. Evol. 34, 3299–3302.
Sahu, M.S., Mali, P.Y., Waikar, S.B., et al., 2010. Evaluation of immunomodulatory potential of ethanolic extract of *Roscoea procera* rhizomes in mice. J. Pharm. BioAllied Sci. 4, 346–349.
Shahzadi, I., Abdullah, Mehmood, F., et al., 2019. Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. Genomics 112, 1454–1463.

Shaw, J., Lichey, E.B., Schilling, E.E., et al., 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am. J. Bot. 94, 275–288.
Small, R.L., Ryburn, J.A., Cronn, R.C., et al., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. Am. J. Bot. 85, 1301–1315.
Srivastava, S., Ankita, M., Kumar, D., et al., 2015. Reversed-phase high-performance liquid chromatography-ultraviolet photodiode array detector validated simultaneous quantification of six bioactive phenolic acids in *Roscoea purpurea* tubers and their in vitro cytotoxic potential against various cell lines. Phcog. Mag. 11, 488–495.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Tang, H.Q., Tang, L., Shao, S.C., et al., 2021. Chloroplast genomic diversity in *Bulbophyllum* section *Macrocaulia* (Orchidaceae, Epidendroideae, Malaxideae): insights into species divergence and adaptive evolution. Plant Divers. 43, 350–361.
Thiel, T., Michalek, W., Varsheny, R., et al., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106, 411–422.
Wei, R., Yang, J., He, L.J., et al., 2021. Plastid phylogenomics provides novel insights into the infrafamilial relationship of Polypodiaceae. Cladistics 37, 717–727.
Wiche, S., Schneeweiss, G.M., dePamphilis, C.W., 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76, 273–297.
Wick, R.R., Schultz, M.B., Zobel, J., et al., 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31, 3350–3352.
Wu, D.L., Larsen, K., 2000. *Roscoea* Sm. In: Wu, Z.Y., Raven, P.H. (Eds.), Flora of China, vol. 24. Science Press, Beijing; Missouri Botanical Garden Press, St. Louis, pp. 362–366.
Wu, Q., Jiang, M., Chen, H.M., et al., 2020. Comparative analysis of three complete chloroplast genomes of *Inula* genus with phylogenetic analysis of 49 plants from Carduoideae. Acta Pharm. Sin. 55, 1042–1049.
Xu, Y.L., Shen, H.H., Du, X.Y., et al., 2022. Plastome characteristics and species identification of Chinese medicinal wintergreens (*Gaultheria*, Ericaceae). Plant Divers. 44, 519–529.
Yang, B.B., Li, L.D., Liu, J.Q., et al., 2021. Plastome and phylogenetic relationship of the woody buckwheat *Fagopyrum tibeticum* in the Qinghai-Tibet Plateau. Plant Divers. 43, 198–205.
Yang, J.B., Li, D.Z., Li, H.T., 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol. Ecol. Resour. 14, 1024–1031.
Yang, J.P., Zhu, Z.L., Fan, Y.J., et al., 2020. Comparative plastomic analysis of three *Bulbophyllum* medicinal plants and its utility for species identification. Acta Pharm. Sin. 55, 2736–2745.
Zhang, D.Q., Duan, L.Z., Zhou, N., 2014. Application of DNA barcoding in *Roscoea* (Zingiberaceae) and a primary discussion on taxonomic status of *Roscoea cautleoides* var. *pubescens*. Biochem. Systemat. Ecol. 52, 14–19.
Zhang, D.Q., Yang, Y.S., Zhou, N., 2015. The identify of *Roscoea cauteoides* var. *pubescens* (Z.Y. Zhu) T.L. Wu based on molecular and morphological evidences. Phytotaxa 205, 259–267.
Zhang, L., Huang, Y.W., Huang, J.L., et al., 2022a. DNA barcoding of *Cymbidium* by genome skimming: call for next-generation nuclear barcodes. Mol. Ecol. Resour. 00, 1–16.
Zhang, Y.M., Han, L.J., Yang, C.W., 2022b. Comparative chloroplast genome analysis of medicinally important *Veratrum* (Melanthiaceae) in China: insights into genomic characterization and phylogenetic relationships. Plant Divers. 44, 70–82.
Zhao, J.L., Xia, Y.M., Cannon, C.H., et al., 2016a. Evolutionary diversification of alpine ginger reflects the early uplift of the Himalayan-Tibetan Plateau and rapid extrusion of Indochina. Gondwana Res. 32, 232–241.
Zhao, J.L., Gugger, P.F., Xia, Y.M., et al., 2016b. Ecological divergence of two closely related *Roscoea* species associated with late Quaternary climate change. J. Biogeogr. 43, 1990–2001.
Zhao, J.L., Zhong, J.S., Fan, Y.L., et al., 2017. A preliminary species-level phylogeny of the alpine ginger *Roscoea*: implications to speciation. J. Syst. Evol. 55, 215–224.
Zhao, J.L., Paudel, B.R., Yu, X.Q., et al., 2021. Speciation along the elevation gradient: divergence of *Roscoea* species within the south slope of the Himalayas. Mol. Phylogenet. Evol. 164, 107292.
Zheng, D.S., Zhang, J.Y., Guo, Q.S., 2013. CpDNA non-coding sequence analysis of *Pinellia ternata* and its related species. Chin. Tradit. Herb. Drugs 44, 881–886.