BMC Plant Biology

RESEARCH

Open Access

Comparative chloroplast genomes of *Incarvillea* species (Bignoniaceae) unveiled genomic diversity and shed light on phylogenetic relationships

Yunhui Jiang[1†], Hong Li[1†], Mei Wu[1], Xuemei Zhang[1], Shukherdorj Baasanmukh[2], Hongzhe Li[1*], Hang Sun[3*] and Shaotian Chen[1*]

Abstract

Background Incarvillea Juss. is a small herbaceous genus within the Bignoniaceae family. It comprises 16 species, which are subdivided into five subgenera. The species are distributed mainly in the Himalaya-Hengduan Mountains, although there are exceptions, including *I. sinensis*, *I. algae*, *I. semiretschenskia*, and *I. potaninii*. Phylogenetic analyses based on *trnL-F* and nr ITS sequences provided support for the monophyly of the genus and its subgenera. However, the interspecific relationships among subgenera remain unresolved, and further investigation is necessary to elucidate these relationships. In this study, we sequenced and assembled 34 chloroplast genomes from 12 *Incarvillea* species, representing all five subgenera, and explored the phylogeny of the genus based on the cp. genome data.

Results The results demonstrated that 34 newly assembled chloroplast genomes exhibited lengths between 159,132 and 169,244 bp, and encoded a total of 129–141 genes. These included 84–95 protein-coding genes, 37 or 38 tRNA genes, and eight rRNA genes. A comparative analysis of the chloroplast genomes revealed the structural rearrangements and the expansions/contractions of the IR regions among the *Incarvillea* species. A total of 12 mutation hotspot regions were identified in the cp. genomes of the genus *Incarvillea*, encompassing six genes (*atpI*, *psaI*, *rps18*, *trnQ-UUG*, *infA* and *ycf1*) and six intergenic spacer regions (*psbT-psbf1*, *rps11-rpl36*, *infA-rps8*, *trnN-GUU-ycf1*, *ndhE-ndhG* and *ndhI-ndhA*). The *Pi* values of all highly variable regions exceeded 0.06. The phylogenetic analysis corroborated the monophyly of the genus and elucidated the relationships between five subgenere, namely ((*Niedzwedzkia*, *Incarvillea*), ((*Amphicome*, *Olgaea*),*Pteroscleris*)).

[†]Yunhui Jiang and Hong Li contributed equally to this work.

*Correspondence:
Hongzhe Li
lihongzheyn@163.com
Hang Sun
sunhang@mail.kib.ac.cn
Shaotian Chen
chenst@mail.kib.ac.cn

Full list of author information is available at the end of the article

Conclusion A comprehensive comparison of cp. genomic sequences revealed the diversity of the genus *Incarvillea* in terms of size, gene content and gene order of cp. genomes. Based on the cp. genome data, a robust phylogenetic tree of the genus *Incarvillea* was generated through phylogenetic analysis, with interspecific relationships well resolved. The results of this study enhance the understanding of the evolutionary history of the genus, and will facilitate further studies on the diversity and resource protection of the genus.

Keywords *Incarvillea*, Chloroplast genome, Phylogenomics, Himalaya-Hengduan mountain, Bignoniaceae

Background

The genus *Incarvillea* Juss. is a small, herbaceous member of the Bignoniaceae family, which is predominantly tropical and woody in nature [1, 2]. It comprises 16 species classified into five subgenera: *Niedzwedzkia* (B. Fedtsch.) Grierson, *Amphicome* (Royle) R. Rr. apud Royle, *Incarvillea* Juss., *Pteroscleris* Baillon, and *Olgaea* S.T. Chen, K.Y. Guan &Z.K. Zhou [1, 3]. These species are primarily distributed across the Himalaya-Hengduan Mountain range, with a few exceptions such as *I. sinensis* Lamarck (extending from Western China to the Russian Far East), *I. algae* Regel and *I. semiretschenskia* Grierson (native to Central Asia), and *I. potaninii* Batalin (native to Mongolia) [2, 3]. *Incarvillea* species are widely recognized for their ethnobotanical and horticultural importance. In southwestern China, *Incarvillea* species, such as *I. arguta* (Royle) Royle, *I. younghusbandii* Sprague, and *I. compacta* Maxim are widely used in traditional medicine for their anti-inflammatory, antibacterial, and antioxidant properties [4–6]. Additionally, their large, colorful flowers enhance their ornamental value. Despite their significance, field surveys in China reveal that many species face severe threats in their natural habitats. Notably, two species, *I. forrestii* Fletcher and *I. altissima* Forrest, are considered extinct due to overgrazing [7], underscoring the urgent need for phylogenetic and genetic diversity studies to inform conservation strategies.

Previous phylogenetic and biogeographical studies based on *trnL-F* sequences and nrITS markers have confirmed the monophyly of *Incarvillea* and identified five distinct subclades, corresponding to its subgenera. However, discrepancies between phylogenetic trees derived from these markers highlight unresolved relationships among subgenera [8, 9]. Furthermore, recent molecular analyses suggest rapid radiation within the subgenus *Pteroscleris*, complicating efforts to resolve its interspecific relationships [8]. The morphological distinctions among the five subgenera, including variations in growth habit, leaf morphology, stamen structure, capsule texture, and seed characteristics, further challenge taxonomic classification and phylogenetic inference [8, 9]. Therefore, a more comprehensive approach integrating additional molecular data is necessary to clarify the phylogenetic relationships and taxonomic status of *Incarvillea*.

In contrast to short DNA markers, chloroplast (cp.) genomes, which encompass a larger number of loci, have become increasingly utilized for reconstructing phylogenetic relationships in angiosperms [10–13]. Most angiosperm cp. genomes exhibit a stable quadripartite structure with maternal inheritance [12, 14, 15]. These genomes, typically ranging from 120 to 160 kb in length, consist of two inverted repeats (IRs) separated by a large single-copy region (LSC) and a small single-copy region (SSC), and generally encode between 120 and 150 genes [16]. Due to their slow mutation rates and lack of recombination, cp. genomes have been extensively applied in phylogenetic and evolutionary studies, as well as species identification within angiosperms [11–13, 17–19]. Comparative analyses of cp. genomes from five *Incarvillea* species have revealed significant differences in genome size, structure, and gene content, along with notable structural rearrangements among the subgenera [20]. In light of these findings, we sequenced the cp. genomes of *Incarvillea* species and conducted a comparative genomic analysis to elucidate their genome structure, composition, genetic diversity, and interspecific phylogenetic relationships. Furthermore, we integrated these cp. genome data with previously reported plastomes from other members of the Bignoniaceae family to clarify the phylogenetic placement of *Incarvillea* within the family.

Results

Structure and characteristics of the *Incarvillea* chloroplast genome

The assembled chloroplast genomes (cp. genomes) of 34 individuals from 12 *Incarvillea* species ranged in size from 159,132 bp (*I. mairei*3) to 169,244 bp (*I. beresovskii*1), with the GC content varying from 39.2% (*I. semiretschenskia*) to 40.4% (*I. mairei*) (Table 1; Fig. 1, Fig. S1). All cp. genomes exhibited a typical quadripartite structure, consisting of a pair of inverted repeats (IRs; 34,129–44,555 bp), separated by a large single-copy region (LSC; 79,741–82,759 bp) and a small single-copy region (SSC; 393–10,111 bp). Notably, the GC content in the IR regions (40.8–42.5%) was higher than in the LSC (37.9–39.3%) and SSC regions (31.3–34.9%). The cp. genomes of *Incarvillea* species contained between 129 and 141 genes, including 84–95 protein-coding genes, 37 or 38 tRNA genes, and 8 rRNA genes (Table 1, Table S1). Of the total number of genes, 18–29 were located in the IR regions and exhibited two copies.

Table 1 Characteristics of the 34 newly assembled Cp. Genomes

species	length(bp)				Gene number (unigene number)				GC content(%)				GenBank Accession No.
	Total	LSC	IR	SSC	Total	CDS	tRNA	rRNA	Total	LSC	SSC	IR	
I. semiretschenskia	163,962	81,723	36,819	8,601	132(112)	87(78)	37(30)	8(4)	39.2	38	33.6	41.2	PQ553075
I. olgae	163,464	82,507	35,913	9,131	132(112)	87(78)	37(30)	8(4)	39.3	37.9	33.8	41.5	PQ553074
*I. arguta*1	161,065	82,244	34,355	10,111	132(111)	87(77)	37(30)	8(4)	39.4	38	34.3	41.9	PQ553054
*I. arguta*2	160,830	82,467	34,129	10,105	130(112)	87(78)	37(30)	8(4)	39.4	38	34.2	42	PQ553055
*I.arguta*3	160,920	82,223	34,301	10,095	132(110)	87(76)	37(30)	8(4)	39.4	38	34.2	41.9	PQ553056
*I. sinensis*1	162,104	82,489	35,602	8,411	132(112)	87(78)	37(30)	8(4)	39.3	38.1	33.3	41.5	PQ553076
*I. sinensis*2	162,486	82,489	35,793	8,411	132(112)	87(78)	37(30)	8(4)	39.3	38.1	33.3	41.5	PQ553077
*I. sinensis*3	162,235	82,506	35,659	8,411	132(112)	87(78)	37(30)	8(4)	39.3	38	33.3	41.5	PQ553078
*I. sinensis*4	162,309	82,710	35,467	8,665	132(112)	87(78)	37(30)	8(4)	39.3	38	33.2	41.6	PQ553079
*I. sinensis*5	162,017	82,315	35,518	8,666	132(111)	87(78)	37(29)	8(4)	39.3	38	33.2	41.6	PQ553080
*I. sinensis*6	162,046	82,759	35,439	8,409	131(112)	86(78)	37(30)	8(4)	39.3	38	33.4	41.6	PQ553081
*I. compacta*1	159,635	80,208	35,146	9,135	132(112)	87(78)	37(30)	8(4)	40.3	39.2	34.8	42.3	PQ553060
*I. compacta*2	159,689	80,250	35,152	9,135	132(112)	87(78)	37(30)	8(4)	40.3	39.2	34.8	42.3	PQ553061
*I. compacta*3	160,787	80,623	35,534	9,096	132(112)	87(78)	37(30)	8(4)	40.3	39.2	34.8	42.2	PQ553062
*I. mairei*1	159,151	79,773	35,145	9,088	132(112)	87(78)	37(30)	8(4)	40.4	39.3	34.9	42.5	PQ553071
*I. mairei*2	160,040	79,997	35,477	9,089	132(112)	87(78)	37(30)	8(4)	40.4	39.3	34.9	42.3	PQ553072
*I. mairei*3	159,132	79,744	35,150	9,088	132(112)	87(78)	37(30)	8(4)	40.4	39.3	34.9	42.4	PQ553073
*I. younghusbandii*1	160,403	80,402	35,538	8,925	130(109)	85(75)	37(30)	8(4)	40.2	39.1	34.7	42.1	PQ553082
*I. younghusbandii*2	160,488	80,449	35,557	8,925	129(109)	84(75)	37(30)	8(4)	40.2	39.1	34.7	42.1	PQ553083
*I. younghusbandii*3	160,522	80,469	35,564	8,925	129(109)	84(75)	37(30)	8(4)	40.2	39.1	34.7	42.1	PQ553084
*I. delavayi*1	165,568	80,560	41,441	2,126	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553063
*I. delavayi*2	165,011	80,381	41,252	2,126	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553064
*I. dissectifoliola*1	164,956	80,700	41,069	2,118	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553065
*I. dissectifoliola*2	164,768	80,615	41,015	2,123	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553066
*I. dissectifoliola*3	164,937	80,615	41,102	2,118	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553067
*I. lutea*1	164,558	80,655	40,890	2,123	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553068
*I. lutea*2	164,855	80,778	40,977	2,123	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553069
*I. lutea*3	164,812	80,735	40,977	2,123	139(111)	93(77)	38(30)	8(4)	40.2	39.2	33.9	41.3	PQ553070
*I. zhongdianensis*1	165,942	81,037	41,395	2,115	138(111)	92(77)	38(30)	8(4)	40.2	39.2	34	41.3	PQ553085
*I. zhongdianensis*2	165,592	80,861	41,308	2,115	139(111)	93(77)	38(30)	8(4)	40.2	39.2	34	41.3	PQ553086
*I. zhongdianensis*3	165,837	80,998	41,362	2,115	138(111)	92(77)	38(30)	8(4)	40.2	39.2	34	41.3	PQ553087
*I. beresovskii*1	169,244	79,741	44,555	393	140(112)	94(78)	38(30)	8(4)	40	39.1	31.3	40.8	PQ553057
*I. beresovskii*2	167,836	79,741	43,851	393	141(112)	95(78)	38(30)	8(4)	40	39.1	31.3	40.9	PQ553058
*I. beresovskii*3	167,836	79,995	43,724	393	138(111)	92(77)	38(30)	8(4)	40	39.1	31.3	40.9	PQ553059

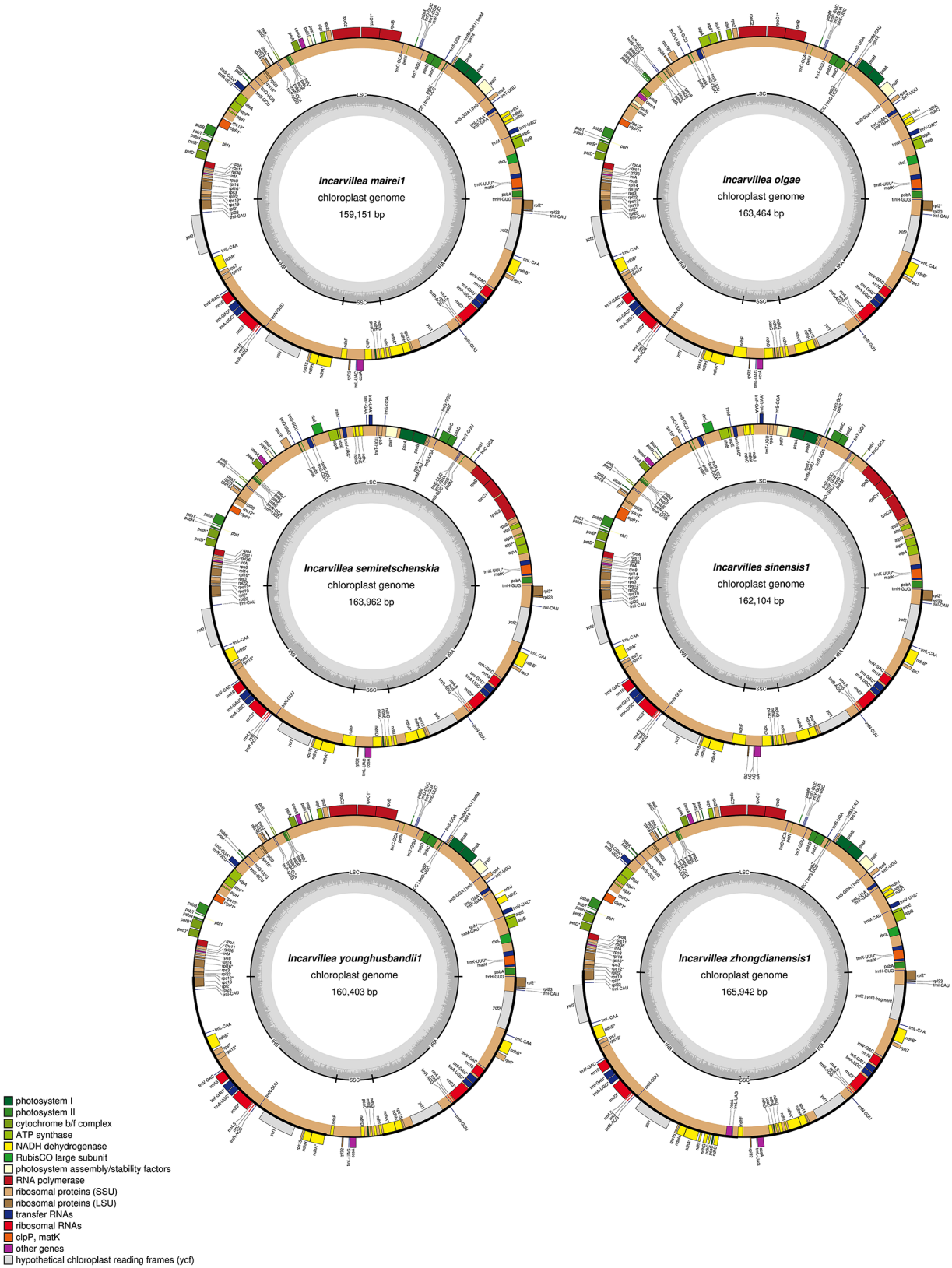

Fig. 1 Chloroplast genome map of 12 *Incarvillea* species. The outer circle shows the genes at each locus, and transcribed clockwise genes are shown outside, while counterclockwise genes are inside. The inner circle indicates the range of the large single-copy (LSC), small single-copy (SSC), and the inverted repeats (IRs). The darker gray area in the inner circle represents the GC content, while the lighter gray corresponds to AT content

IR contraction and expansion

Comparative analysis of the IR boundaries revealed considerable variation in the expansion and contraction of the IR regions across the 34 *Incarvillea* genomes (Fig. 2). This variation contributed to differences in the total cp. genome sizes and IR region lengths among species. Within the subgenus *Pteroscleris*, a strong positive correlation (r^2 = 0.9841) was observed between the size of the cp. genomes (159,132–169,244 bp) and the length of the IR regions (35,145–44,555 bp) (Table 1, Fig. S2).

The LSC/IRb boundary was consistently located within the *rps19* gene, extending into IRb by 14 bp in all species. The LSC/IRa boundary was positioned between *trnH* and *rpl2* across all sequenced species. However, the SSC/IR boundary exhibited notable variability among species. For the junction sites of SSC and IRb, three distinct configurations were identified: (1) In four species (*I. arguta*, *I. compacta*, *I. mairei*, and *I. younghusbandii*), the SSC/IRb boundary was located between the *ndhA* and *ndhF* genes; (2) Seven species (*I. dissectifoliola*, *I. lutea*, *I. zhongdianensis*, *I. delavayi*, *I. sinensis*, *I. semiretschenskia*, and *I. olgae*) had the SSC/IRb boundary within the *ndhF* gene, with the gene extending into IRb by 51 to 571 bp; (3) In *I. beresovskii*, the SSC/IRb boundary was located 784 bp from *ndhF*, and no genes were present in the SSC region. Furthermore, this species exhibited the same junction structure for SSC/IRa as that for SSC/IRb. For the SSC/IRa boundary, the remaining 11 *Incarvillea* species (excluding *I. beresovskii*) exhibited three configurations: (1) In five species (*I. compacta*, *I. mairei*, *I. sinensis*, *I. semiretschenskia*, and *I. olgae*), the boundary was situated between *ndhL* and *ndhA*; (2) In two species (*I. arguta* and *I. younghusbandii*), the boundary was within the *ndhL* gene, extending into IRa by 1–10 bp; (3) In four species (*I. dissectifoliola*, *I. lutea*, *I. zhongdianensis*, and *I. delavayi*), the boundary was located within the *rps32* gene, extending into IRa by 51–52 bp.

Repeat structure and SSR analysis of chloroplast genomes

The cp. genomes of *Incarvillea* species contained between 192 and 696 long repetitive sequences, including 85–358 palindromic (P) repeats and 107–338 forward (F) repeats, but no reverse (R) or complementary (C) repeats were identified (Fig. 3, Table S2). The majority of long repeats were concentrated in the IR regions, with fewer located in the SSC region. The longest palindromic and forward repeats were 270 bp in length, detected in *I. delavayi*, *I. dissectifoliola*, and *I. zhongdianensis*.

A total of 1489 simple sequence repeats (SSRs) were identified across the 12 *Incarvillea* species (Fig. 4, Table S3). The majority of SSRs were located in the LSC region, with mononucleotide SSRs comprising 60.11% of the total. These monomeric SSRs were predominantly A/T repeats, indicating a base composition bias towards A/T. The number of SSRs per species ranged from 34 (*I. arguta*2) to 52 (*I. sinensis*2). Notably, six species (*I. beresovskii*, *I. compacta*, *I. delavayi*, *I. dissectifoliola*, *I. lutea*, and *I. mairei*) lacked dinucleotide repeats, while hexanucleotide SSRs were absent in *I. arguta*, *I. beresovskii*, and *I. semiretschenskia*.

Divergence in *Incarvillea* cp. genomes and hypervariable regions

Synteny analysis using the Mauve software revealed considerable rearrangements in the LSC regions of the cp. genomes across *Incarvillea* species (Fig. 5). However, nine species, including *I. arguta*, *I. compacta*, *I. mairei*, *I. younghusbandii*, *I. dissectifoliola*, *I. lutea*, *I. zhongdianensis*, *I. delavayi*, and *I. beresovskii*, exhibited relatively conserved cp. genome structures. In contrast, notable genomic rearrangements were observed between this group and the remaining *Incarvillea* species.

Sequence identity analysis with mVISTA indicated a higher degree of variation in non-coding regions compared to coding regions (Fig. 6). The average nucleotide diversity (Pi) was significantly higher in non-coding regions (0.0276) than in coding regions (0.0187) (Fig. 7, Table S2). Sliding window analysis identified 12 hypervariable regions with Pi values exceeding 0.06, including six genes (*atpI*, *psaI*, *rps18*, *trnQ-UUG*, *infA*, and *ycf1*) and six intergenic spacer regions (*psbT-psbf1*, *rps11-rpl36*, *infA-rps8*, *trnN-GUU-ycf1*, *ndhE-ndhG*, and *ndhI-ndhA*) (Fig. 7). Eight mutation hotspots were located in the LSC region, while three hypervariable regions (*trnQ-UUG*, *infA*, ycf1) were in the IR region, and the final region (*ndhI-ndhA*) was near the SSC/IRa boundary.

Phylogenetic relationships

Based on 69 common CDS, the trees constructed using the maximum likelihood method and the Bayesian method exhibited a high degree of similarity in topology, with the exception of the position of *Catalpa bignonioides* Walter (Fig. 8, Fig. S3). Phylogenetic analysis revealed strong support for the monophyly of *Incarvillea* (BP = 100, PP = 1) and its close relationship with *Tecomaria capensis* (Thunb.) Spach (BP = 100, PP = 1). The subgenus *Pteroscleris* formed a distinct clade with full support (BP = 100, PP = 1), and was sister to a clade comprising *I. olgae* and *I. arguta*. *I. semiretschenskia* and *I. sinensis* were clustered into a clade with full support, forming a sister group to the remaining species in the genus. A distinct phylogenetic clade was observed within *I. sinensis*, which clustered all six accessions into a single group, with full support. This clade was further divided into two sister groups: one containing two annual individuals and the other comprising four perennial individuals.

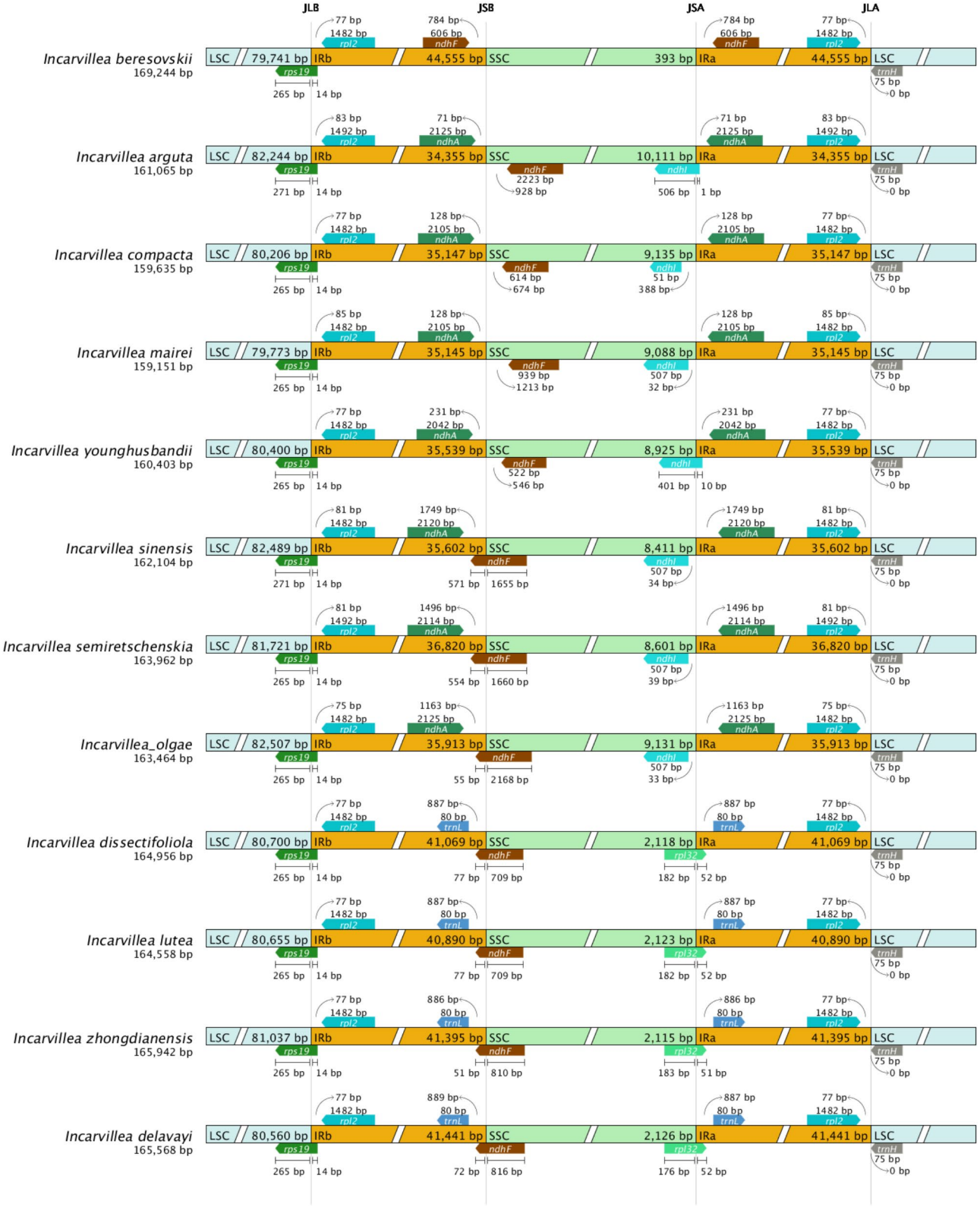

Fig. 2 Comparison of the junctions among LSC, SSC and IR regions within 12 cp. genomes of the *Incarvillea* species

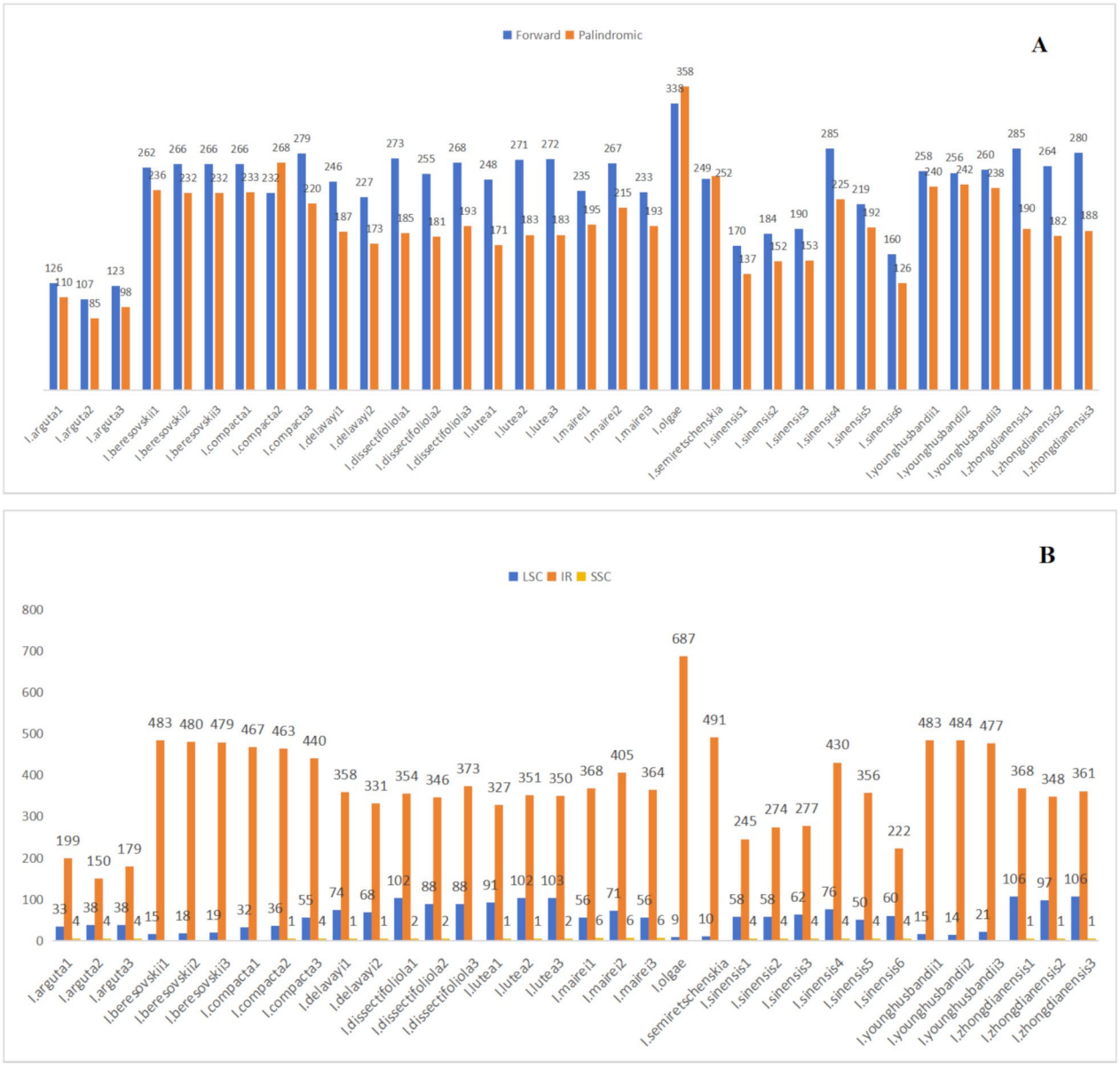

Fig. 3 Statistics of tandem repeats in 34 cp. genome sequences of 12 *Incarvillea* species. (**A**) The type and number of tandem repeats; (**B**) Distribution of tandem repeats

Phylogenetic relationships inferred from six coding genes showed consistent results with the 69 CDS-based phylogeny (Fig. S4). In contrast, the phylogeny based on the IGS regions revealed significant divergence from the 69 CDS-based tree (Fig. S5).

Discussion

Diversity of Chloroplast genome characteristics

In this study, we obtained chloroplast genomes for 34 individuals from 12 species of *Incarvillea*, representing all five subgenera: the monotypic *Niedzwedzkia* and *Olgaea*, the binary *Amphicome* and *Incarvillea*, and the polytypic *Pteroscleris*. To ensure accuracy in the assembly and annotation of the chloroplast genomes, we sampled 2 to 3 individuals per species, except for *I. semiretschenskia* and *I. olgae*. Our findings showed a high level of concordance in chloroplast genome architecture among intraspecific individuals, despite minor insertions or deletions between individuals within species.

The expansion and contraction of inverted repeat (IR) regions are recognized as significant evolutionary phenomena within chloroplast (cp.) genomes, influencing the variations in chloroplast genome size [21–24]. We observed substantial expansion and contraction of IR

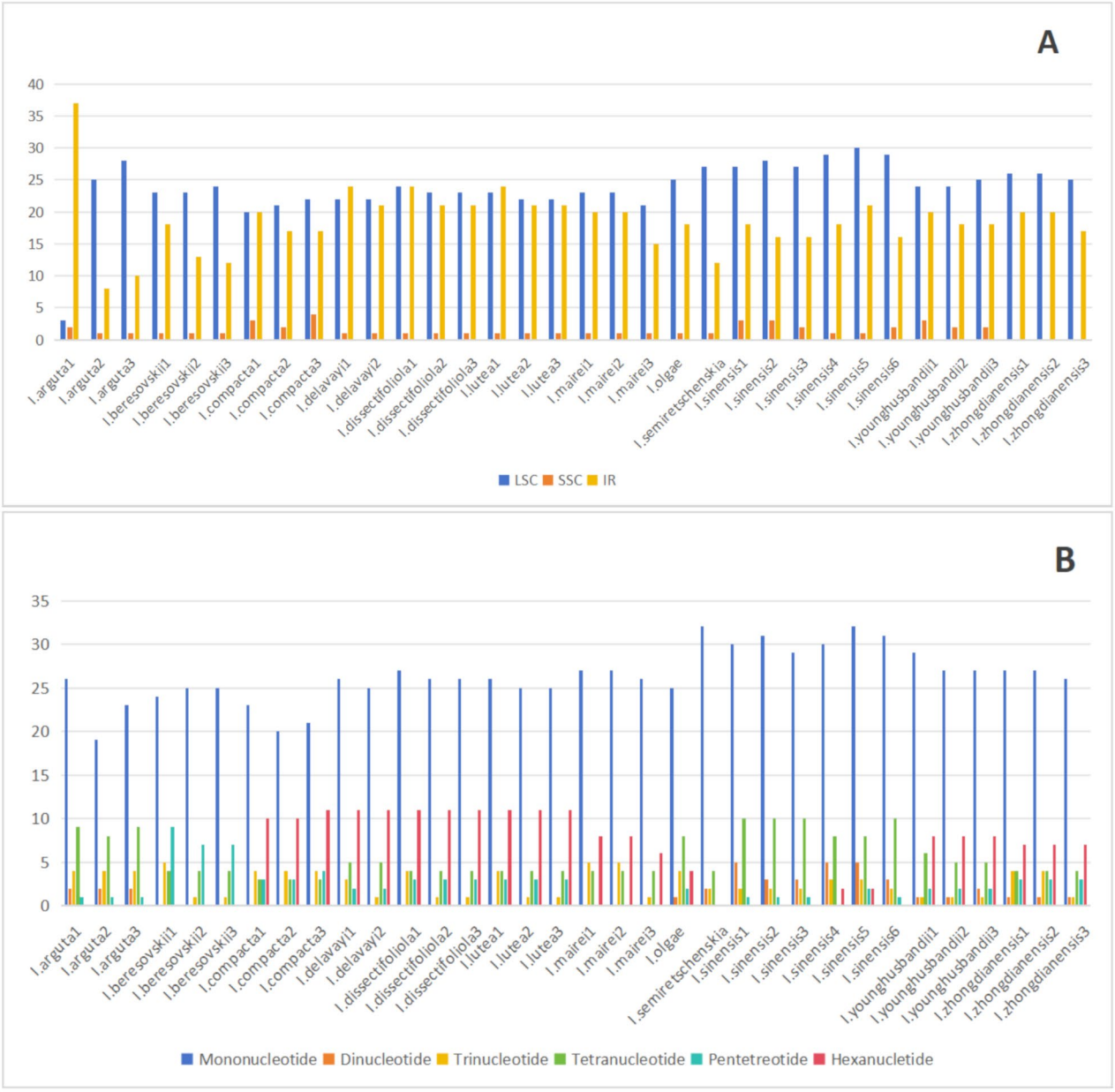

Fig. 4 Simple sequence repeats in 34 cp. genomes of 12 *Incarvillea* species. (**A**) Distribution of SSRs; (**B**) The type and number of SSR

regions in the cp. genomes of *Incarvillea* species. Additionally, we found a positive correlation between the chloroplast genome size and the length of the IR regions in the *Pteroscleris* subgenus. Previous studies have indicated that closely related species tend to exhibit similar responses to environmental changes, suggesting that the observed shifts in IR boundaries may be non-critical [25, 26]. Interestingly, species in the *Pteroscleris* subgenus showed little genetic variation and share high morphological similarity in previous studies [8, 9], yet they exhibited notable expansion and contraction of the IR regions, leading to significant variations in chloroplast genome size and shifts in the IR boundary (Fig. 2; Table 1). In contrast, the other four subgenera exhibited smaller chloroplast genome size variations and more modest IR boundary fluctuations. These four monotypic or binary subgenera demonstrated pronounced genetic divergence and distinct morphological traits [8, 9]. We hypothesize that the observed discrepancies among subgenera may be linked to significant lineage extinction events associated with these four monotypic or binary subgenera.

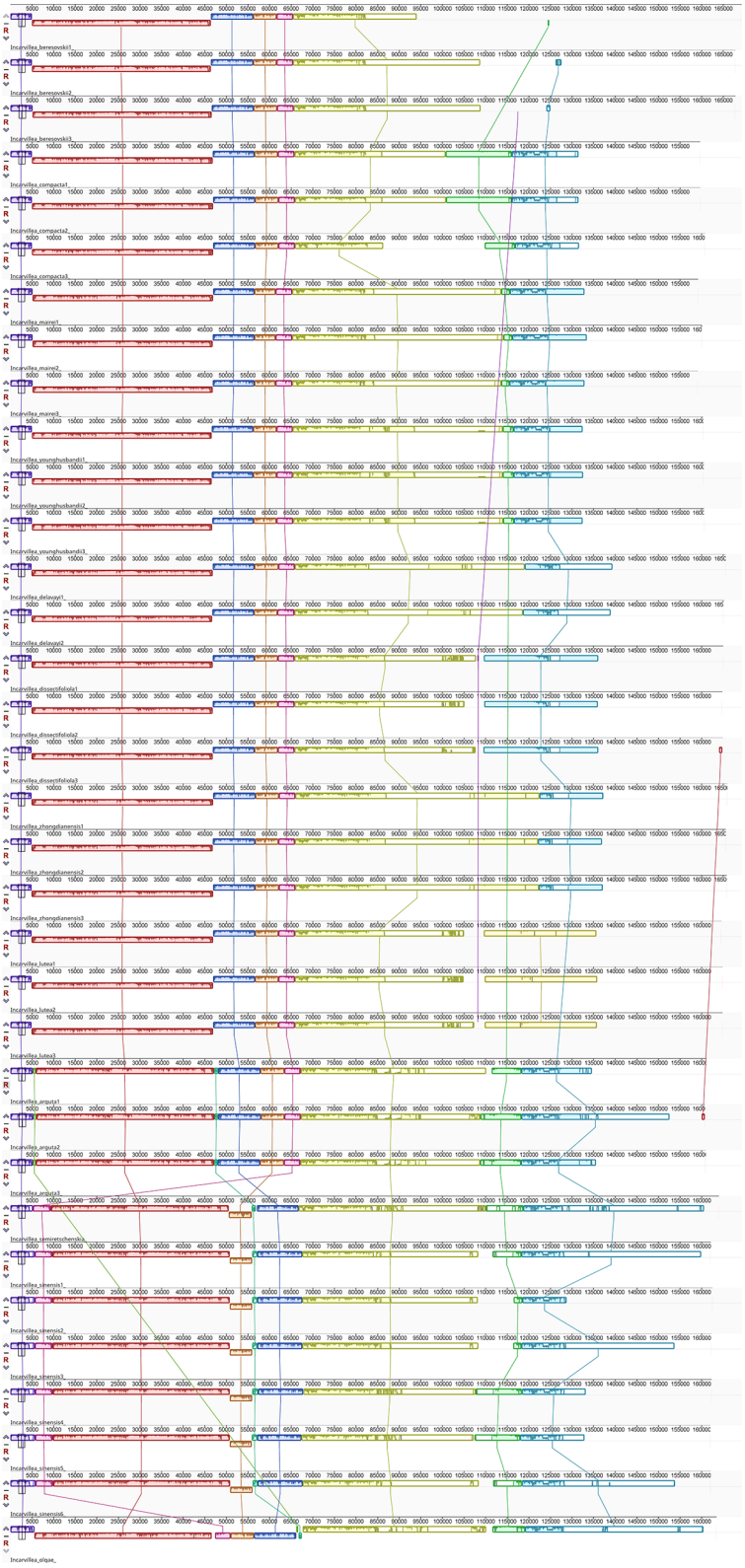

Fig. 5 Synteny analysis of 34 chloroplast genomes from 12 *Incarvillea* species

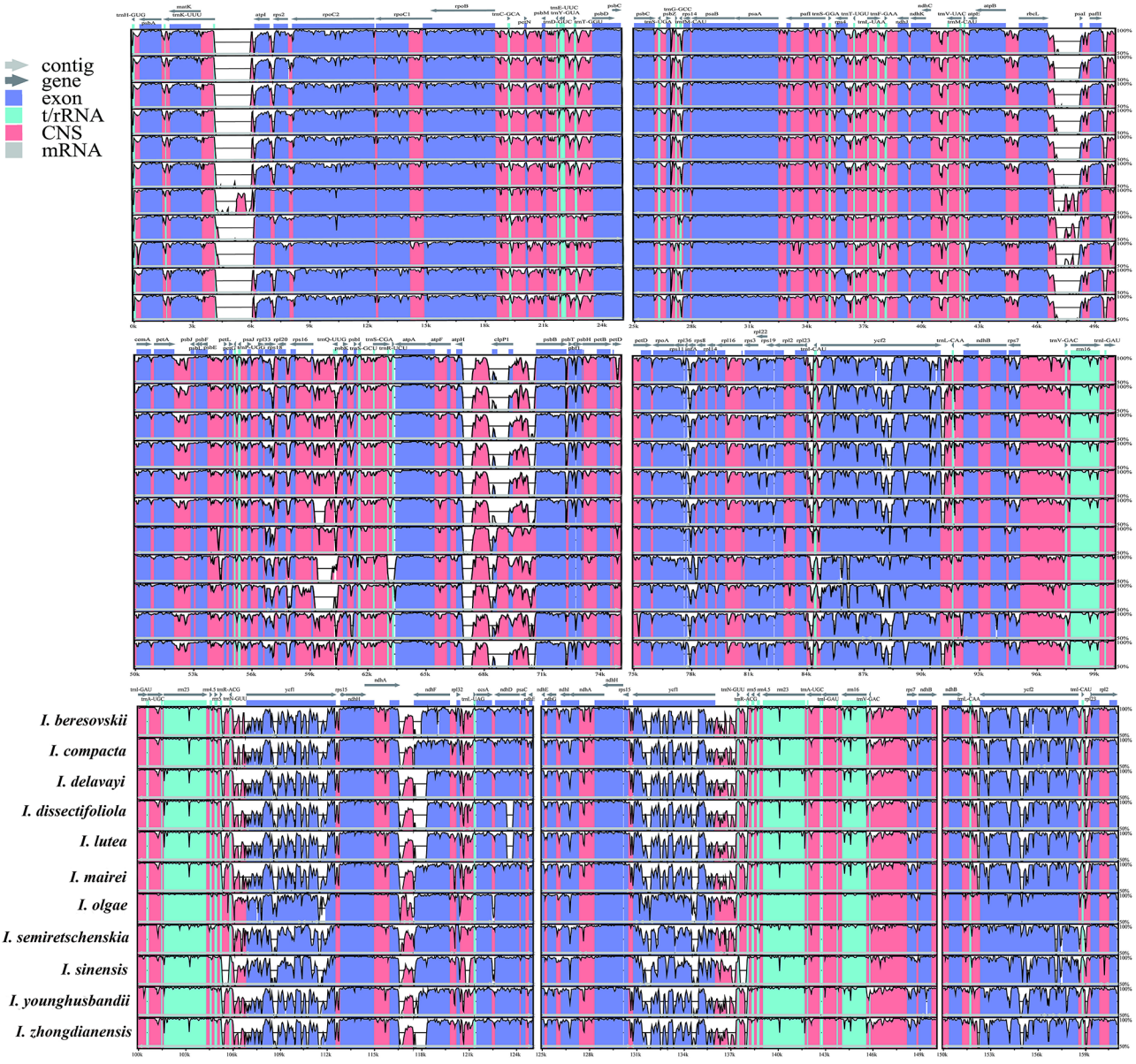

Fig. 6 12 sequences alignment was performed by mVISTA using *Incarvillea arguta* as a reference. The vertical scale represents the percentage of identity, ranging from 50 to 100%

Phylogenetic relationships among subgenera of *Incarvillea*

The genus *Incarvillea* comprises 16 species, classified into five subgenera [1, 3]. These subgenera are distinguished based on karyomorphological and morphological characteristics, including chromosome size, growth habit, anther texture, capsule shape, and seed wing structure [8, 9, 27]. This diversity complicates the intuitive classification of these subgenera within the same genus. Historically, the *Amphicome* subgenus was classified as a distinct genus within Gesneriaceae or Bignoniaceae [28–30]. Similarly, *Niedzwedzkia* was originally described as a genus in the Pedaliaceae family and later assigned to the Bignoniaceae as an independent genus [31, 32]. Molecular phylogenetic analyses using ITS and *trnL-F* fragments positioned the genus within Bignoniaceae, but phylogenetic trees based on ITS or *trnL-F* sequences have often failed to provide strong support, leading to occasional conflicts regarding the relationships among subgenera [8, 9].

Chloroplast genome sequences have proven valuable for elucidating phylogenetic relationships among angiosperms [33]. However, due to significant rearrangements in gene order, complete cp. genomes was unsuitable for direct phylogenetic reconstruction within *Incarvillea*. We therefore conducted phylogenetic inference using shared coding sequences (CDS) among *Incarvillea* species and

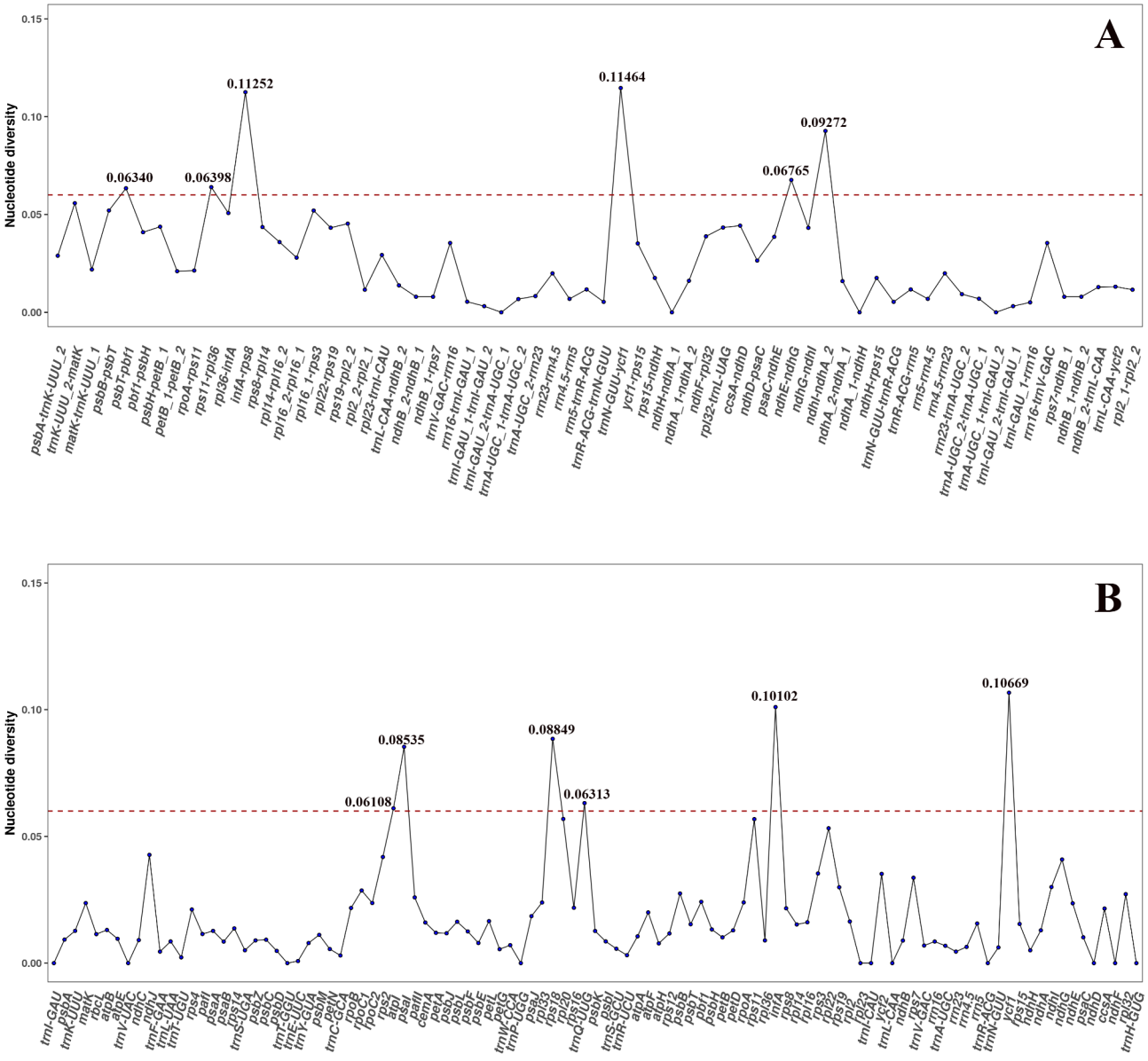

Fig. 7 The graph of nucleotide diversity (Pi) values of intergenic spacer regions (**A**) and genes (**B**)

related taxa. Our analysis confirmed the monophyletic status of *Incarvillea*, and further supported its placement within the family Bignoniaceae. The phylogenetic relationships among the five recognized subgenera were well-supported, with *Pteroscleris* being sister to the clade containing *Olgae* and *Amphicome*. The remaining subgenera, *Incarvillea* and *Niedzwedzkia*, formed a distinct clade, which was intriguing as it represented a sister group to a large clade containing the other three subgenera. The phylogenetic analysis based on the *trnL-F* sequence fragment showed that the subgenus *Niedzwedzkia* was the sister group to the clade formed by the subgenus *Incarvillea* and the subgenus *Pteroscleris* [9]. However, this relationship was supported very weakly, and the relationship of this clade with other clades was unresolved, which seemed to imply the instability in the topological structure of the phylogenetic tree based on the *trnL-F* sequence. Our study presents a robust phylogenetic tree with high resolution and strong support for the relationships among subgenera and species.

Lineage divergence in subgenus *Pteroscleris*

The *Pteroscleris* subgenus is geographically confined to the Himalayan-Hengduan Mountain region. Within this subgenus, divergences of interspecific morphological traits are predominantly quantitative, and the continuous variation and overlap in these traits complicate species delineation and identification [1, 2]. Molecular

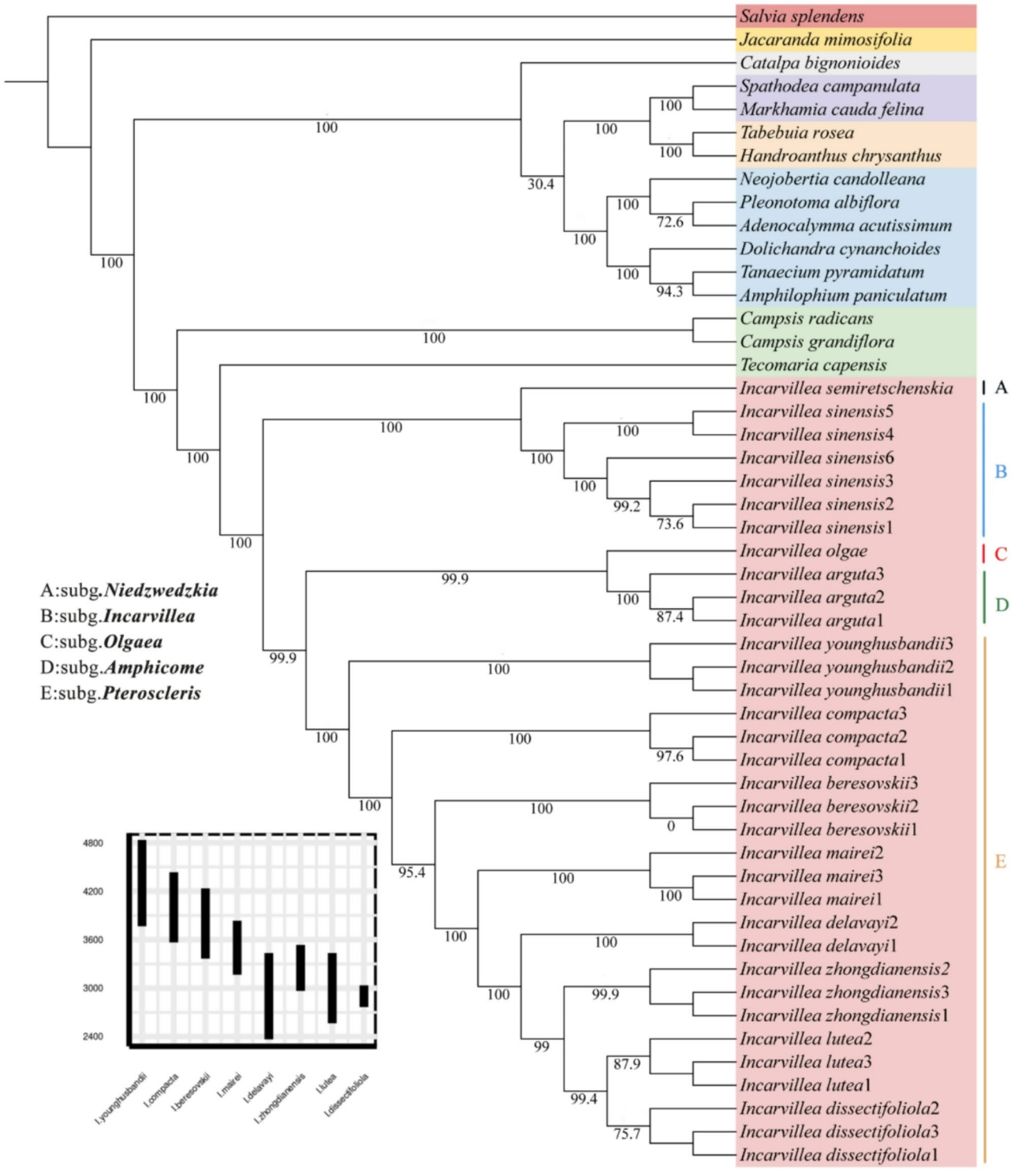

Fig. 8 The Phylogenetic tree constructed using Maximum likelihood based on 69 common CDS. Numbers below branches indicate Maximum Likelihood bootstrap support. The elevation range of the distribution areas for each species of the *Pterocleris* subgenus is displayed in the lower left corner of the image. The lower left corner shows the altitudinal range of the distribution area for species of the subgenus *Pteroscleris*

phylogenetic studies suggest that the *Pteroscleris* subgenus has undergone a recent radiation, with speciation events closely associated with the rapid uplift of the Himalayan Mountains, and interspecific relationships are ambiguous [8]. Our chloroplast genome-based phylogenetic analysis provided compelling evidence for the interspecific evolutionary relationships within this subgenus, supporting *I. younghusbandii* as the basal species, followed by *I. compacta*. This result is consistent with previous ITS-based phylogenetic analyses [8, 9]. Geographically, these two species are located at the highest altitudes in the subgenus, while other species found at lower elevations occupy deeper nodes in the phylogeny (Fig. 8). We propose that the dramatic uplift of the Himalayas may have induced spatial isolation between these two species and other members of the subgenus, facilitating their evolutionary divergence.

Divergence of *Incarvillea sinensis*

Incarvillea sinensis is the most widely distributed species in the genus, ranging from Western China to Northeastern China and the Russian Far East [1, 2]. Populations distributed in the Himalaya-Hengduan Mountains and their adjacent areas are perennial, while populations in other areas are annual. Phylogeographic studies indicated potential genetic divergence between the annual and perennial populations of *I. sinensis* [34]. Furthermore, reproductive isolation between the two groups has been observed, leading to the hypothesis that the annual and perennial forms of *I. sinensis* may represent separate species [35]. In this study, we sampled six individuals, including two annual and four perennial specimens. Our results of the phylogenetic analysis indicated that they diverged into two distinct branches with full support, indirectly supporting the hypothesis that these groups may represent separate species. Thus, *Incarvillea sinensis* presents an excellent model for investigating how geographic vicariance drives speciation.

Potential molecular markers for *Incarvillea*

The structure of chloroplast genomes is generally conservative among angiosperm species, but mutation hotspots can be identified even among closely related species [36, 37]. Therefore, the chloroplast genome provides a rich source of molecular markers for phylogenetic studies, species identification, and biodiversity assessments. A sliding window analysis of 34 cp. genomes from *Incarvillea* revealed 12 highly variable regions, characterized by Pi values exceeding 0.06. These included six genes (*atpI*, *psaI*, *rps18*, *trnQ-UUG*, *infA*, and *ycf1*) and six intergenic spacer (IGS) regions (*psbT-psbf1*, *rps11-rpl36*, *infA-rps8*, *trnN-GUU-ycf1*, *ndhE-ndhG*, and *ndhI-ndhA*). These regions may serve as valuable molecular markers in future studies on the genetic diversity and species identification of *Incarvillea*.

The maximum likelihood tree constructed based on six coding gene sequences with Pi values greater than 0.06 exhibited consistent topology with that derived from 69 shared CDS (Fig. S4). However, the sequence alignment showed considerable divergence among the six IGS regions with Pi values above 0.06, leading to numerous ambiguous or erroneous regions within the dataset. Consequently, some clades were questionable and difficult to interpret in the IGS-based tree (Fig. S5). We propose that IGS regions evolve more rapidly than coding genes, making them more suitable for intraspecific studies, and care should be taken when IGS regions are used for highly.

Conclusion

This study provided a comprehensive analysis of the chloroplast genomes of 34 individuals representing 12 *Incarvillea* species, shedding light on the structural diversity, genomic characteristics, and evolutionary patterns within the genus. Through comparative genomic analysis, we identified significant variability in chloroplast genome sizes, inverted repeat (IR) region expansions and contractions, and the genome arrangement among Incarvillea species. These findings offer valuable insights into the evolutionary processes shaping *Incarvillea* species. Phylogenetic analysis based on shared coding sequences revealed a robust monophyletic structure for the genus *Incarvillea*, with well-resolved deep nodes, enhancing our understanding of the evolutionary relationships within the group. Importantly, this study identified 12 hypervariable regions in the chloroplast genomes of Incarvillea, including six genes (*atpI*, *psaI*, *rps18*, *trnQ-UUG*, *infA*, and *ycf1*) and six intergenic spacer regions (*psbT-psbf1*, *rps11-rpl36*, *infA-rps8*, *trnN-GUU-ycf1*, *ndhE-ndhG*, and *ndhI-ndhA*). These regions exhibit high nucleotide diversity and hold considerable potential as molecular markers for future phylogenetic research, species identification, and biodiversity assessments within the genus. Overall, the findings of this study significantly contribute to our understanding of the evolutionary history of *Incarvillea*, providing a foundation for further investigations into speciation, evolutionary diversity, and resource conservation within this genus.

Methods

Sampling, DNA extraction, and chloroplast genome sequencing

A total of 34 individuals representing 12 species of *Incarvillea* were sampled for this study (Table S4). Of these, ten species were collected from the field, and fresh, healthy leaves were subsequently dried using silica gel. Two to three individuals were sampled for each species, with the exception of *I. sinensis*, for which six specimens

were collected, including two annual and four perennial individuals. None of these species are endangered or protected, and no specific permits are required for the collection of specimens.Voucher specimens were identified by Shaotian Chen, and have been deposited at the Museum of Ethnic Medicine, Yunnan University of Chinese Medicine. Genomic DNA was extracted from leaf tissue using a modified CTAB method [38]. DNA quality and concentration were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). Libraries with an insert size of 300–500 bp were constructed using the NEBNext Ultra II DNA PCR-free Library Prep Kit according to the manufacturer's instructions. The libraries were then sequenced on the Illumina NovaSeq platform (Illumina Inc., USA) in PE 150 nt mode by Personal Biotechnology Company (Shanghai, China). For the two *Incarvillea* species endemic to Central Asia, only single individuals were available, which were obtained from the National Herbarium of Uzbekistan (TASH). DNA extraction, library construction and sequencing for these species followed the protocol outlined by Zeng et al. [39]. To reconstruct the phylogenetic position of the genus *Incarvillea*, we retrieved the chloroplast genome sequences of 15 species from other genera within the Bignoniaceae family, as well as one species from the Lamiaceae family, from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) (Table S5).

Chloroplast genome assembly and annotation

Raw sequencing data were adapter-trimmed and quality-filtered using AdapterRemoval v2 (trimwindows = 5 and minlength = 50) [40]. De novo assembly of the chloroplast genomes was performed using GetOrganelle v1.7.4 [41], with the chloroplast genome of *I. sinensis* (MT937254) as the reference. The assembled genomes were annotated using the online tool GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) [42] and manually curated in Geneious [43] to correct potential annotation errors. Physical maps of the cp. genomes were generated using OGDRAW v1.3.1 (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) [44]. The annotated chloroplast genome sequences of the 12 *Incarvillea* species have been deposited in the GenBank database (accession numbers PQ553054 - PQ553087).

Analysis of SSRs and repeat sequences

Simple sequence repeats (SSRs) in the newly sequenced chloroplast genomes were identified using the MISA online tool (https://webblast.ipk-gatersleben.de/misa/) [45]. The minimum number of repeats for each type of SSR was set as follows: mononucleotides, 10; dinucleotides, 5; trinucleotides, 4; tetranucleotides, 3; pentanucleotides, 3; and hexanucleotides, 3. Tandem repeat sequences were detected using the REPuter program [46], identifying forward, reverse, palindrome, and complementary sequences. Parameters for repeat detection included a minimum repeat size of 30, a Hamming distance of 3, and a similarity threshold of 90%.

Comparative analysis of chloroplast genome

The IR/SC boundaries and adjacent genes of the chloroplast genomes of *Incarvillea* species were visualized using the CPJSdraw v1.0.0 tool [47] to examine the expansion and contraction of the IR region. The chloroplast genome sizes, GC content, and lengths of the LSC, SSC, and IR regions were calculated using Geneious v2022.0.1 [43]. To assess divergence among the chloroplast genomes of *Incarvillea* species, 34 chloroplast genome sequences were aligned using MAFFT v7.490 [48], and sequence similarity profiles were visualized using the mVISTA tool (https://genome.lbl.gov/vista/index.shtml) [49]. Additionally, a synteny analysis was performed using Mauve [50]. Nucleotide diversity (Pi) values were calculated using CPStools [51], and a sliding window analysis was conducted with a window size of 600 bp and a step size of 200 bp to identify regions of high sequence divergence.

Phylogenetic analysis

To investigate the taxonomic relationships and interspecific relationships within the genus *Incarvillea*, phylogenetic analyses were performed using both maximum likelihood (ML) and Bayesian inference (BI) methods. The analysis was based on the conserved coding sequences (CDS) of chloroplast genomes from 27 species in the Bignoniaceae family, with one species from the Lamiaceae family used as the outgroup, which is the closest related family to Bignoniaceae and is optimal to root the phylogenetic tree. In total, 54 accessions from 32 species were included in the analysis. The common CDS were extracted using PhyloSuite v1.2.3 [52], aligned with MAFFT with default settings [44], and manually adjusted. The ML tree was constructed using IQ-TREE v2.2.0 with 1,000 bootstrap replicates [53] under the automatically selected model. The Bayesian analysis was performed using MrBayes v3.2.7 [54], employing the same evolutionary model as used in the maximum likelihood analysis, with Markov Chain Monte Carlo (MCMC) simulations run for 10 million generations. Sampling occurred every 1,000 generations to ensure adequate exploration of the parameter space. A burn-in of 25% was applied to discard the initial trees, and the remaining trees were used to construct a majority-rule consensus tree, calculating the posterior probability (PP) for each branch.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06380-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Supplementary Material 7

Supplementary Material 8

Acknowledgements
Authors thank Dr. Jiahui Chen and Zhengshan He (Kunming Institute of Botany, CAS) for their assistance in data collection. We also thank Dr. Komiljon Tojibaev and Ziyoviddin Yusupov from Institute of Botany, Academy of Sciences of Uzbekistan for kindly providing sample materials.

Author contributions
Conceptualization, HZ.L., H.S. and S.C.; methodology, Y.J. and H.L.; validation, Y.J. and H.L., formal analysis, Y.J. and H.L., investigation, M.W., X.Z., S.B. and H.Z.L.; resources, M.W., X.Z. and S.B.; data curation, M.W., X.Z. and H.Z.L.; writing—original draft preparation, Y.J., H.L. and S.C.; writing—review and editing, S.B., HZ.L., H.S. and S.C.; visualization, Y.J. and H.L., supervision, H.S.; project administration, HZ.L.; funding acquisition, H.S. and S.C. All authors have read and agreed to the submitted version of the manuscript.

Funding
This work was supported by the Key Projects of the Joint Fund of the National Natural Science Foundation of China (U23A20149) and the Joint Special Project of Yunnan Province Science and Technology Department on Basic Research of Traditional Chinese Medicine (202101AZ070001–160).

Data availability
New sequenced and other published chloroplast genome sequences can be found in the GenBank (https://www.ncbi.nlm.nih.gov/genbank/), and the accession numbers are shown in Table 1 and Table S2.

Declarations

Ethics approval and consent to participate
This study complied with relevant institutional, national, and international guidelines and legislation. This article does not involve any endangered or protected species, and no specific permits are required for the collection of specimens.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Traditional Chinese Medicine, Yunnan University of Chinese Medicine, Kunming 650500, China
[2]Department of Biology & Chemistry, Changwon National University, Changwon, South Korea
[3]Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

Received: 25 January 2025 / Accepted: 10 March 2025
Published online: 29 March 2025

References

1. Grierson AJC. A revision of the genus Incarvillea. Notes R Bot Gard Edinb. 1961;23:303–54.
2. Wang WT, Pan KY, Zhang ZY, Li ZY, Tao DD, Yin WQ. Incarvillea juss. In: Wang WT, editor. Flora reipublicae popularis sinicae. Volume 69. Beijing: Science; 1990. pp. 34–49.
3. Chen ST, Guan KY, Zhou ZK. A new subgenus of Incarvillea (Bignoniaceae). Ann Bot Fenn. 2006;43:288–90.
4. Luo YG, Yi JH, Li BG, Zhang GL. Novel ceramides and a new glucoceramide from the roots of Incarvillea Arguta. Lipids. 2004;39(9):907–13.
5. Pan WG, Jiang SP, Luo P, Gao P, Chen B, Bu HT. Isolation, purification and structure identification of antioxidant compound from the roots of Incarvillea younghusbandii Sprague and its life span prolonging effect in drosophila melanogaster. Nat Prod Res. 2008;22(8):719–25.
6. Shen T, Li XQ, Hu WC, Wang W, Chen L, Bu XT, Wang Y, Wang YJ, Wang XJ, Chen KX, Guo YW, Shao CL, Wang CY. Hepatoprotective effect of phenylethanoid glycosides from Incarvillea compacta against CCl4-induced cytotoxicity in HepG2 cells. J Korean Soc Appl Biol Chem. 2015;58(4):617–25.
7. Chen ST, Gong J, Guan KY, Zhou ZK. Biodiversity conservation of the genus Incarvillea Juss. (Bignoniaceae) based on molecular diversity and species richness assessment. J Plant Biol. 2010;53(6):387–94.
8. Chen ST, Guan KY, Zhou ZK, Olmstead R, Cronk Q. Molecular phylogeny of Incarvillea (Bignoniaceae) based on ITS and trnL-F sequences. Am J Bot. 2005;92(4):625–33.
9. Rana SK, Luo D, Rana HK, Chen S, Sun H. Molecular phylogeny, biogeography and character evolution of the montane genus Incarvillea Juss. (Bignoniaceae). Plant Divers. 2021;43(01):1–14.
10. Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol. 2011;76:273–97.
11. Dong WP, Xu C, Cheng T, Lin K, Zhou S. Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. Genome Biol Evol. 2013;5(5):989–97.
12. Yang Z, Wang GX, Ma Q, Ma WX, Liang LS, Zhao TT. The complete Chloroplast genomes of three betulaceae species: implications for molecular phylogeny and historical biogeography. Peer J. 2019;7:e6320.
13. Wang YH, Wang S, Liu YL, Yuan QL, Sun JH, Guo LP. Chloroplast genome variation and phylogenetic relationships of atractylodes species. BMC Genomics. 2021;22(1):1–12.
14. Palmer JD. Comparative organization of Chloroplast genomes. Annu Rev Genet. 1985;19(1):325–54.
15. Clegg MT, Gaut BS, Learn GH Jr, Morton BR. Rates and patterns of Chloroplast DNA evolution. Proc Natl Acad Sci. 1994;91(15):6795–800.
16. Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK. The complete Chloroplast genome sequence of pelargonium×hortorum: organization and evolution of the largest and most highly rearranged Chloroplast genome of land plants. Mol Biol Evol. 2006;23(11):2175–90.
17. Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuch M, Chang Z, Aota SI, Inokuch H, Ozek H. Chloroplast gene organization deduced from complete sequence of liverwort Marchantia polymorpha Chloroplast DNA. Nature. 1986;322(6079):572–4.
18. Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng BY, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M. The complete nucleotide sequence of the tobacco Chloroplast genome: its gene organization and expression. EMBO J. 1986;5(9):2043–9.
19. Hiratsuka J, Shimada H, Whittier R, Ishibashi K, Sakamoto M, Mori M, Kondo C, Honji Y, Sun CR, Meng BY, Li YQ, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M. The complete sequence of the rice (Oryza sativa) Chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet. 1989;217:185–94.
20. Ma HX, Zhang YY, Zhang YW, Yixi ZM, Liang ZS, Li XR, Guo YT, Feng S. Chloroplast Genome Comparison Phylogenetic Anal Five Med Plants Incarvillea. 2024;55:5972–81.
21. Kim KJ, Lee HL. Complete Chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res. 2004;11:247–61.
22. Davis JI, Soreng RJ. Migration of endpoints of two genes relative to boundaries between regions of the plastid genome in the grass family (Poaceae). Am J Bot. 2010;97:874–92.

23. Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen camellia Chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol. 2014;14:151.
24. Li DM, Li J, Wang DR, Xu YC, Zhu GF. Molecular evolution of Chloroplast genomes in subfamily Zingiberoideae (Zingiberaceae). BMC Plant Biol. 2021;21(1):558.
25. Downie SR, Jansen RK. A comparative analysis of whole plastid genomes from the apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot. 2015;40:336–51.
26. Zhu A, Guo W, Gupta S, Fan W, Mower JP. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 2016;209(4):1747–56.
27. Chen ST, Zhou ZK, Guan KY, Nakata M. Karyomorphological study of Incarvillea Juss. (Bignoniaceae) and its implications in distribution and taxonomy. Bot J Linn Soc. 2004;144:113–21.
28. Don G. A general history of the dichlamydeous plants. J. G. and F. Rivington, London, UK. 1838.
29. Bureau E. Monographie des Bignoniacées: recherches Sur les caractères, La composition, les limites de Cet Ordre naturel de Plantes. L. Martinet, Paris, France. 1864.
30. Burtt BL, Grierson AJC. The name Incarvillea Arguta. Notes Roy Bot Gard Edinb. 1953;19:155–6.
31. Vassilczenko IT, Niedzwedzkia. Not Syst Union Rerumpubl Soc. 1957;18:219–24.
32. Vassilczenko IT, Niedzwedzkia. Flora Union Rerumpubl Soc. 1958;23:5–12.
33. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc Natl Acad Sci. 2007;104(49):19363–8.
34. Chen ST, Xing YW, Su T, Zhou ZK, Dilcher DL, Soltis DE. Phylogeographic analysis reveals significant Spatial genetic structure of Incarvillea sinensis as a product of mountain Building. BMC Plant Biol. 2012;12:58.
35. Lan WJ, Wang FY, Barrett S, Wang WT, Ma Y, Yang Y, Li N, Deng JC. BaiWN. Mechanisms of reproductive isolation between annual and perennial plants of Incarvillea sinensis. J Syst Evol. 2024;62(3):548–60.
36. Scarcelli N, Mariac C, Couvreur TLP, Faye A, Richard D, Sabot F, Berthouly-Salazar C, Vigouroux Y. Intraindividual polymorphism in chloroplasts from NGS data: where does it come from and how to handle it? Mol Ecol Resour. 2016;16:434–45.
37. Xu X, Shen Y, Zhang Y, Li Q, Wang W, Chen L, Chen S, Zhang W, Xu JC, Qiu DY. A comparison of 25 complete Chloroplast genomes between sister Mangrove species Kandelia obovata and Kandelia Candel geographically separated by the South China sea. Front Plant Sci. 2022;13:1075353.
38. Doyle J. DNA protocols for plants. In: Hewitt GM, Johnston AWB, Young JPW, editors. Molecular techniques in taxonomy. Berlin, Heidelberg: Springer; 1991. pp. 283–93.
39. Zeng CX, Hollingsworth PM, Yang J, He ZS, Zhang ZR, Li DZ, Yang JB. Genome skimming herbarium specimens for DNA barcoding and phylogenomics. Plant Methods. 2018;14:1–14.
40. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. 2016;9(1):88.
41. Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. GetOrganelle: a fast and versatile toolkit for accurate de Novo assembly of organelle genomes. Genome Biol. 2020;21(1):241.
42. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
43. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647–9.
44. Greiner S, Lehwark P, Bock R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47(W1):W59–64.
45. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5.
46. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001;29:4633–41.
47. Li H, Guo Q, Xu L, Gao H, Liu L, Zhou X. CPJSdraw: analysis and visualization of junction sites of Chloroplast genomes. PeerJ. 2023;11:e15326.
48. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47(W1):W5–10.
49. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:W273–9.
50. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004;14(7):1394–403.
51. Huang LJ, Yu HX, Wang Z, Xu WB. CPStools: a package for analyzing Chloroplast genome sequences. iMetaOmicse. 2024; 25.
52. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour. 2020;20(1):348–55.
53. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268–74.
54. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012;61(3):539–42.

Publisher's note