

RESEARCH

Open Access



Phylogeny and diversification of genus *Sanicula* L. (Apiaceae): novel insights from plastid phylogenomic analyses

Bo-Ni Song¹, Chang-Kun Liu¹, An-Qi Zhao¹, Rong-Ming Tian¹, Deng-Feng Xie¹, Yu-Lin Xiao¹, Huai Chen², Song-Dong Zhou^{1*} and Xing-Jin He^{1*}

Abstract

Background The genus *Sanicula* L. is a unique perennial herb that holds important medicinal values. Although the previous studies on *Sanicula* provided us with a good research basis, its taxonomic system and interspecific relationships have not been satisfactorily resolved, especially for those endemic to China. Moreover, the evolutionary history of this genus also remains inadequately understood. The plastid genomes possessing highly conserved structure and limited evolutionary rate have proved to be an effective tool for studying plant phylogeny and evolution.

Results In the current study, we newly sequenced and assembled fifteen *Sanicula* complete plastomes. Combined with two previously reported plastomes, we performed comprehensively plastid phylogenomics analyses to gain novel insights into the evolutionary history of this genus. The comparative results indicated that the seventeen plastomes exhibited a high degree of conservation and similarity in terms of their structure, size, GC content, gene order, IR borders, codon bias patterns and SSRs profiles. Such as all of them displayed a typical quadripartite structure, including a large single copy region (LSC: 85,074–86,197 bp), a small single copy region (SSC: 17,047–17,132 bp) separated by a pair of inverted repeat regions (IRs: 26,176–26,334 bp). And the seventeen plastomes had similar IR boundaries and the adjacent genes were identical. The *rps19* gene was located at the junction of the LSC/IRa, the IRa/SSC junction region was located between the *trnN* gene and *ndhF* gene, the *ycf1* gene appeared in the SSC/IRb junction and the IRb/LSC boundary was located between *rpl12* gene and *trnH* gene. Twelve specific mutation hotspots (*atpF*, *cemA*, *accD*, *rpl22*, *rbcl*, *matK*, *ycf1*, *trnH-psbA*, *ycf4-cemA*, *rbcl-accD*, *trnE-trnT* and *trnG-trnR*) were identified that can serve as potential DNA barcodes for species identification within the genus *Sanicula*. Furthermore, the plastomes data and Internal Transcribed Spacer (ITS) sequences were performed to reconstruct the phylogeny of *Sanicula*. Although the tree topologies of them were incongruent, both provided strong evidence supporting the monophyly of Saniculoideae and Apioideae. In addition, the sister groups between Saniculoideae and Apioideae were strongly suggested. The *Sanicula* species involved in this study were clustered into a clade, and the *Eryngium* species were also clustered together. However, it was clearly observed that the sections of *Sanicula* involved in the current study were not respectively recovered as monophyletic group. Molecular dating analysis explored that the origin of this genus was occurred during the late Eocene period, approximately 37.84 Ma (95% HPD: 20.33–52.21 Ma) years ago and the diversification of the genus was occurred in early Miocene 18.38 Ma (95% HPD: 10.68–25.28 Ma).

*Correspondence:

Song-Dong Zhou

zsd@scu.edu.cn

Xing-Jin He

xjhe@scu.edu.cn

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusion The plastome-based tree and ITS-based tree generated incongruences, which may be attributed to the event of hybridization/introgression, incomplete lineage sorting (ILS) and chloroplast capture. Our study highlighted the power of plastome data to significantly improve the phylogenetic supports and resolutions, and to efficiently explore the evolutionary history of this genus. Molecular dating analysis explored that the diversification of the genus occurred in the early Miocene, which was largely influenced by the prevalence of the East Asian monsoon and the uplift of the Hengduan Mountains (HDM). In summary, our study provides novel insights into the plastome evolution, phylogenetic relationships, taxonomic framework and evolution of genus *Sanicula*.

Keywords Apiaceae, *Sanicula*, Plastome, DNA barcodes, Phylogeny, Divergence time

Background

Sanicula L. is a unique genus of perennial woodland herbs belonging to Apiaceae subfamily Saniculoideae [1]. The genus consists of about forty species worldwide that are widely distributed in Asia, North America, Europe and Africa [1–4], with two distribution centers: China and North America. The relatively widespread distribution of this genus is mainly due to the hooked prickles on the fruit that can stick to animals, which facilitate the spread of these species to further areas [5].

The genus, known as black snakeroot or sanicle, was established by Linnaeus in 1753 with the type species *S. europaea* L [1, 6]. The main morphological features of its members are characterized by spiny or bristly simple leaves, simple (rarely compound) umbels or heads, showy bracts, conspicuously prominent and persistent calyx, two persistent styles and fruits covered with scales, bristles, or hooked prickles, which can easily distinguish them from other genera of the Apiaceae [1, 7]. However, the genus always exhibits great variability in habit, foliage, flowers and fruits [8], thus some researchers were prone to divide the genus into more smaller classification units. For example, De Candolle separated the genus into two sections: *Sanicla* DC. and *Sanicoria* DC. [9, 10] relied on leaf and fruit morphology. However, Drude divided the genus into three subgenera (*Sanicula* L., *Sanicoria* DC. and *Erythrosana* Drude) and seven sections (*Rosiflorae* Drude, *Flaviflorae* Drude, *Orthospermae* Drude, *Camplospermae* Drude, *Pinnatae* Drude, *Tuberculatae* Drude and *Tuberosae* Drude) based on the petal color, seed surface shape and fruit prickles [11]. On the basis of Drude, Wolff divided the genus into ten sections [12]. Later, Shan and Constance proposed a completely revised treatment for this genus, who reduced the four sections of Wolff's classification to the synonymy taxonomy in sect. *Sanicoria* DC and formed five sections (*Tuberculatae* Drude, *Pseudopetagnia* H.wolff, *Sanicula* DC., *Sandwicenses* Shan and Constance, *Sanicoria* DC.) based on differences in habit, flower, fruit and distribution [8]. Therefore, it can be seen that the classification system of genus *Sanicula* has long been disputed. In addition, species delimitation of this genus was also

blurred, which was also mainly caused by the high variability of morphological characteristics. For example, *S. orthacantha* var. *stolonifera* Shan & S.L.Liou has long rhizomes with distinct nodes, so it was considered to be a variety of *S. orthacantha* S. Moore [13]. Li et al. observed that *S. pengshuiensis* M. L. Sheh & Z. Y. Liu was morphologically similar to *S. lamelligera* Wolff ex Kretsch, and thus treated it as a synonymy of the latter [14]. From previous studies, it can be seen that the traditional methods of distinguishing these species were mainly focused on their morphological features, whereas many species of the genus have always shown great morphological instability, which have lead to extreme difficulties in species classification and delimitation [8–12]. Therefore, more evidence are needed to re-evaluate the taxonomic system of the genus.

A robust phylogenetic framework can provide strong evidence for the taxonomy of the genus *Sanicula*. Hence, some molecular markers have been applied to the phylogenetic studies of this genus, including single or multiple-locus DNA sequence data (for example ITS sequences, plastid DNA *rpl16*, *rps16* intron and *rpoC1* intron, *trrnQ-trn K* 5'-exon). Among these studies, Vargas P et al. observed that the sect. *Sanicoria* was not a monophyletic group and suggested that the circumscription of sect. *Sanicoria* should be enlarged to include the species of sect. *Sandwicenses* [5]. Valiejo et al. subsequently elucidated the relationships among the main genera of Saniculoidea based on nrITS sequences and found that sec. *Pseudopetagnia* and sect. *Sanicula* were also not monophyletic group [15]. In addition, there was much controversy about the interspecific relationships of the genus. For example, Vargas P et al. thought *S. orthacantha* was more closely related to *S. lamelligera* [5], while Yang et al. considered *S. orthacantha* had affinity to *S. chinese* [16]. However, these phylogenetic trees generated from DNA fragments had weak supports and low resolutions, and failed to provide sufficient information to support the improvement of taxonomy for *Sanicula* [5, 7, 17–19]. The sectional taxonomy system and the interspecific relationships of genus *Sanicula* are facing a severe challenge. Therefore, additional molecular resources are urgently

needed to reconstruct a strongly robust phylogeny of genus *Sanicula* and re-evaluate the sectional taxonomy and interspecific relationships of the genus.

Besides, the evolutionary history of *Sanicula* is still poorly understood. Kadereit et al. proposed that the genus originated in the Miocene period of the Tertiary (15.5–6.0 Ma) based on the ITS sequences and plastid fragment *rps16*, combined with molecular clock [19]. Wen et al. indicated the origin of the genus was occurred at 16.54 Ma (95% HPD: 8.06–27.67 Ma) referred from 79 CDS data [20]. Vargas P et al. estimated the divergence time of sect. *Sanicoria* and sect. *Sandwicenses* relied on nrDNA sequences [5]. Although previous studies have laid the foundation for exploring the evolutionary history of the genus, the results obtained may be unreliable due to limited sampling and DNA fragments resulting in low supports and resolutions of the phylogenetic trees. Thus, a reliable phylogenetic framework and time tree are urgently needed to further investigate the evolutionary history of the genus based on expanded sampling.

Additionally, many species of *Sanicula* are valuable traditional herbs [21]. For example, *S. lamelligera*, known as “Fei Jingcao”, is a well-known traditional Chinese medicine that is widely used to treat cold, cough, tracheitis, bruises and amenorrhea [13, 22]. Other species such as *S. orthacantha* S. Moore., *S. orthacantha* var. *brevispina* de Boiss., *S. caerulescens* Franch., *S. chinensis* Bunge., *S. elata* Hamilt., and *S. astantiifolia* Wolff ex Kretsch. are also Chinese medicinal herbs with the effects of resolving phlegm and cough, activating blood circulation and removing blood stasis [13]. Nevertheless, misuse of species names occurred frequently due to the abundant morphological variations within species, such as *S. chinensis* and *S. orthacantha*, as well as *S. caerulescens* and *S. lamelligera* [23], which made it difficult to identify species accurately. Hence, it is necessary to develop more efficient and specific DNA barcodes for *Sanicula* species authentication to ensure medicinal quality.

Plastid is a crucial and special multifunctional organelle involved in processes such as photosynthesis, carbon fixation and numerous major biological metabolic processes [24, 25]. The plastid genome (plastome) of flower plants ranges in size from 120 kb to 170 kb and displays a typical quadripartite structure, comprising a large single copy (LSC: 82–90 kb), a small single copy (SSC: 15–20 kb) and a pair of inverted repeats (IRa & IRb: 22–25 kb) [26]. It always encodes 110–130 unique genes, including about 80 protein-coding genes (PCGs), 30 transfer RNA (tRNAs) genes and four ribosomal RNA (rRNAs) genes [27]. The plastome is highly conserved in terms of its gene structure, sequence, content, order compared to nuclear and mitochondrial genomes [28, 29]. Moreover, the plastome is uniparental inheritance

and possesses highly variable sites [30]. For these obvious advantages, the plastome has become a powerful tool to generate highly supported and resolved phylogenies and to explore more efficient specific DNA barcodes [25, 27, 28, 31–33]. In recent years, with the rapid development of next-generation sequencing technology and bioinformatics technology, a large number of sequence data of plastomes has become more accessible at a much lower cost [34], and it has been extensively and successfully applied to solve the plants phylogenies, especially for those taxonomically difficult taxa within the family Apiaceae [30–32, 35–40].

Plastomes also opened a promising window for disclosure of the genus *Sanicula*. Currently, although there were related studies on the plastomes of genus *Sanicula*, all of them only reported the *Sanicula* plastome structure or conducted a simple analyses, which failed to address the phylogenetic relationships of the genus [16, 41–43]. In the current study, we obtained seventeen plastomes of *Sanicula* species (fifteen newly sequenced and two previously reported) and performed comprehensively plastid phylogenomics analyses of this taxonomically difficult genus. Since the plastid genome is maternally inherited, it does not fully reveal the evolutionary history of the taxon, while ITS sequence is biparentally inherited and represents another different genetic pattern [30]. Therefore, to obtain a comprehensive understanding of phylogenetic relationships, we also used ITS sequence to construct the phylogenetic tree. Our major aims were to: (1) characterize the plastomes of *Sanicula* plants and select highly variable hotspot regions as candidate DNA barcodes for species authentication of *Sanicula*; (2) evaluate the potential of plastome for resolving the phylogeny of genus *Sanicula* and re-evaluate the section-level classification system and interspecific relationships; (3) investigate the evolutionary history of the genus.

Results

Basic characteristics of *Sanicula* plastomes

The complete plastome size of the seventeen *Sanicula* taxa ranged from 154,500 bp (*S. odorata* (Raf.) Pryer & Phillippe.) to 155,847 bp (*S. lamelligera*) (Table 1). All of them exhibited a typical quadripartite structure, including a large single copy region (LSC: 85,074–86,197 bp), a small single copy region (SSC: 17,047–17,132 bp) separated by a pair of inverted repeat regions (IRs: 26,176–26,334 bp) (Fig. 1; Table 1). The total GC content was account for 38.20%, except for *S. astantiifolia* (38.10%); the IR regions had the highest GC content (42.9% – 43.0%) and the SSC region had the lowest GC content (32.4%– 32.6%). Four rRNAs (*rrn16*, *rrn23*, *rrn4.5* and *rrn5*) genes possessed the same GC content (55.3%) among seventeen plastomes. All plastomes encoded 113

Table 1 Comparison of plastome features among seventeen *Sanicula* plants

Taxon	Length (bp)				GC contents (%)				Number of genes (unique)			
	Genome	LSC	SSC	IR	Total	LSC	SSC	IR	Total	PCGs	rRNA	tRNA
<i>S. astantiifolia</i>	155,748	86,140	17,096	26,256	38.10	36.4	32.5	42.9	113	79	4	30
<i>S. chinensis</i>	155,395	85,677	17,072	26,322	38.20	36.5	32.6	42.9	113	79	4	30
<i>S. caerulescens</i>	155,768	86,174	17,118	26,238	38.20	36.4	32.6	42.9	113	79	4	30
<i>S. elongata</i>	155,634	86,051	17,105	26,239	38.20	36.4	32.5	42.9	113	79	4	30
<i>S. giraldii</i>	155,609	85,857	17,086	26,333	38.20	36.5	32.5	42.9	113	79	4	30
<i>S. hacquetioides</i>	155,494	85,771	17,075	26,324	38.20	36.5	32.5	42.9	113	79	4	30
<i>S. lamelligera</i>	155,847	86,197	17,132	26,259	38.20	36.4	32.5	42.9	113	79	4	30
<i>S. orthacantha</i>	155,749	86,166	17,103	26,240	38.20	36.4	32.6	42.9	113	79	4	30
<i>S. orthacantha</i> var. <i>stolonifera</i>	155,396	85,818	17,098	26,240	38.20	36.4	32.5	42.9	113	79	4	30
<i>S. oviformis</i>	155,645	86,071	17,098	26,238	38.20	36.4	32.6	42.9	113	79	4	30
<i>S. pengshuiensis</i>	155,765	86,189	17,102	26,237	38.20	36.4	32.6	42.9	113	79	4	30
<i>S. rubriflora</i>	155,657	85,982	17,057	26,308	38.20	36.4	32.5	42.9	113	79	4	30
<i>S. rugulosa</i>	155,527	85,780	17,079	26,334	38.20	36.5	32.5	42.9	113	79	4	30
<i>S. serrata</i>	155,599	86,046	17,097	26,228	38.20	36.5	32.4	42.9	113	79	4	30
<i>S. tienmuensis</i>	155,717	86,075	17,106	26,268	38.20	36.4	32.5	42.9	113	79	4	30
<i>S. odorata</i>	154,500	85,074	17,074	26,176	38.20	36.5	32.5	42.9	113	79	4	30
<i>S. flavovirens</i>	155,335	85,682	17,047	26,217	38.20	36.4	32.5	43.0	113	79	4	30

unique genes, including 79 PCGs, 30 tRNA genes and 4 rRNA genes (Table 1) and these unique genes had four categories: Self-replication, Genes for photosynthesis, Other genes and Genes of unknown function. Among them, 46 coding genes were related to photosynthesis, which were divided into six group: Subunits of photosystem I, Subunits of photosystem II, Subunits of cytochrome, Subunits of ATP synthase, Large subunit of Rubisco and Subunits of NADH dehydrogenase. Seventeen genes contained intron and fourteen of them (*ndhB*, *trnA*, *ndhA*, *trnI*, *rpl2*, *rpl16*, *petB*, *trnV*, *trnL*, *rpoC1*, *atpF*, *trnG*, *rps16* and *trnK*) harbored one intron and three of them (*clpP*, *rps12* and *ycf3*) harbored two introns (Table S1).

Codon usage bias and relative synonymous codon usage (RSCU) analysis

In order to analyze the codon usage of *Sanicula* plastomes, we screened the CDSs with the length more than 300 bp and finally attained 53 CDSs. We identified 21,162–21,202 codons in these seventeen plastomes, which have similar codon usage bias. The Leu had the highest number of codons (2,210–2,240) and it was encoded by UUA, UUG, CUU, CUC, CUA and CUG. Ser was the second most abundant amino and it was encoded by AUU, AUC and AUA, while the Cys had the least codons (217–265) with UGU and UGC encoded except for TER (Table S2). Among these 64 codons, the codon AUU occurred the most and the frequency was reached

to 850–874, while UGC was the least (51–82 occurrences) except three terminator codons (Fig. 2, Table S2). The most frequent terminator codons was UAA with 29–52 occurrences, while the other two were UAG (15–17) and UGA (14–49), respectively. The RSCU values of the 64 codons unchanged significantly and showed similar codon preferences with values ranging from 0.34 to 1.91 (Fig. 2). Notably, the RSCU values for thirty codons was greater than 1.00 and six codons (UUA, UCU, ACU, GCU, UAU and AGA) were regarded as “overrepresented” codons with RSCU values greater than 1.6 except for GAU. The remaining 32 codons had RSCU values less than 1, and fifteen of them (CUC, CUG, GUC, AGC, ACG, GCG, UAC, CAG, AAC, GAC, GAG, UGC, UGC, CGG and GGC) were considered to be “underrepresented” codons with RSCU values less than 0.6. Two codons (GCU and GUC) displayed no preferences (RSCU=1). All codons with RSCU > 1.00 ended with an A or a U, except for UUG codon (Fig. 2).

Repeat sequences and simple sequence repeats (SSRs) analysis

In this study, we detected 782 repeat sequences in seventeen plastomes and classified them into four types: forward repeats, palindromic repeats, reverse repeats, and complementary repeats. Among them, the most frequent repeats was the palindromic repeats (390), compared to other three types of repeats: forward repeats (376), reverse repeats (13) and complementary repeats (3)

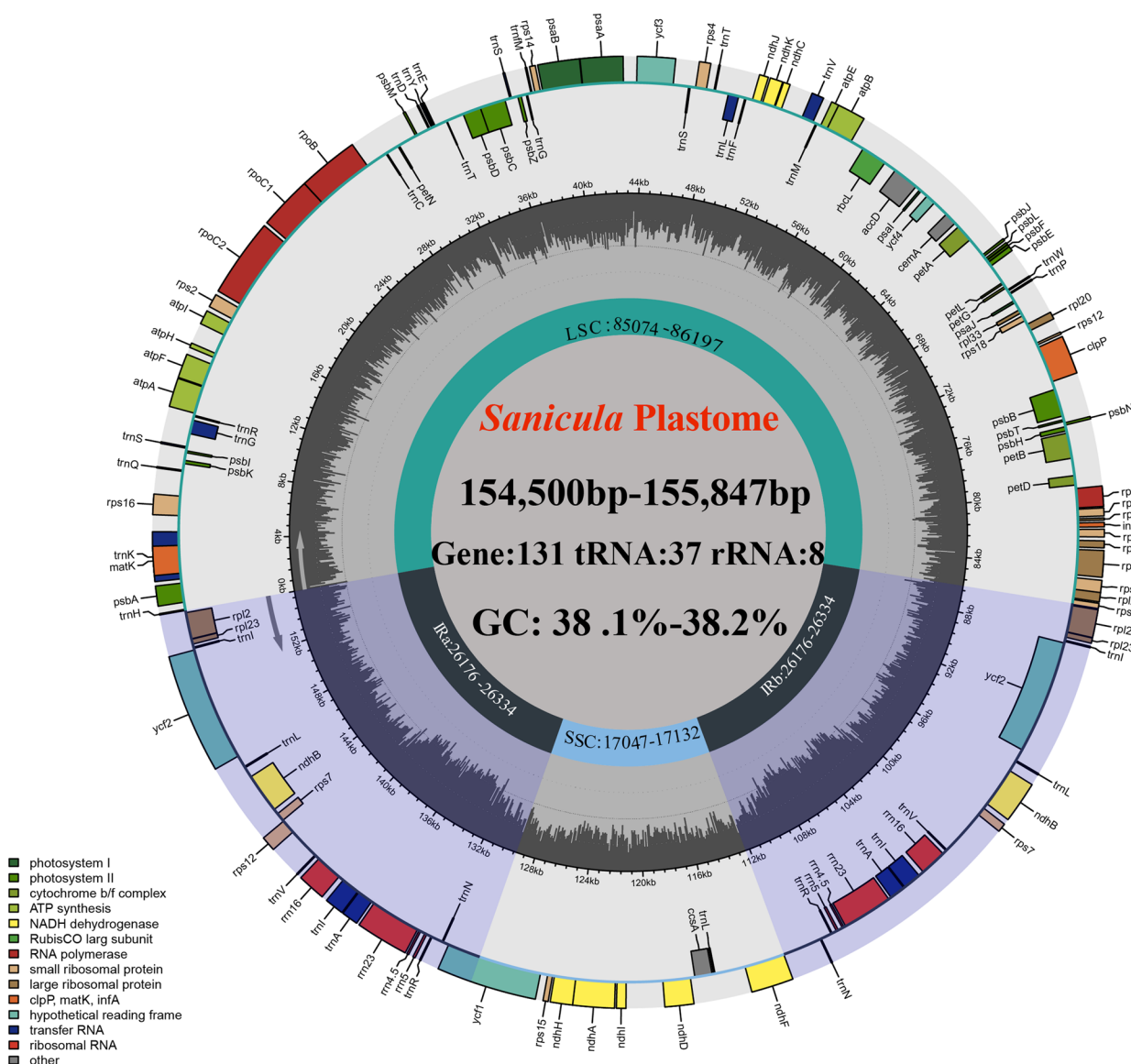


Fig. 1 Whole plastome maps of seventeen *Sanicula* species. Genes showed outside of outward layer circle are transcribed clockwise, while those insides are transcribed counterclockwise. The genes belonging to different functional groups are color-coded. The pale gray area of the inner circle denotes the GC content of plastome

(Fig. 3A, Table S3). The species – *S. rugulosa* Diels. – had the least repeats (42), while three species (*S. lamelligera*, *S. serrata* H. Wolff., *S. tienmuensis* Shan & Constance.) had the most repeats (49). Forward and palindromic repeats were present in seventeen plastomes, and the complementary repeats only occurred in *S. lamelligera* and *S. odorata*. Nine *Sanicula* plastomes had reverse repeats. (Fig. 3A, Table S3).

In addition, we identified 979 SSRs in seventeen plastomes (Table S3). Among these SSRs, the mono-repeats were the most abundant (460, 46.99%), followed by

di-repeats 365 (37.28%), tetra-repeats 111 (11.34%), tri-repeats 38 (3.88%), hexanucleotide repeats 3 (0.31%) and penta-repeats were 2 (0.2%). All plastomes had the mono-repeats, di-repeats, tri-repeats and tetra-repeats (Fig. 3B). The penta-repeats and hexanucleotides were very rare, with penta-repeats (AAAAT/ATTTT and ACTAT/AGTAT) appearing only in *S. flavovirens* Z.H.Chen, D.D. Ma et W. Y. Xie. and hexanucleotide (ACATAT/ATATGT) occurring only in *S. hacquetioides* Franch., *S. rugulosa* and *S. serrata*. In the mono-repeats, A/T motifs were more abundant (24–32) than G/C motif (0–2). Likewise, AT sequences were also

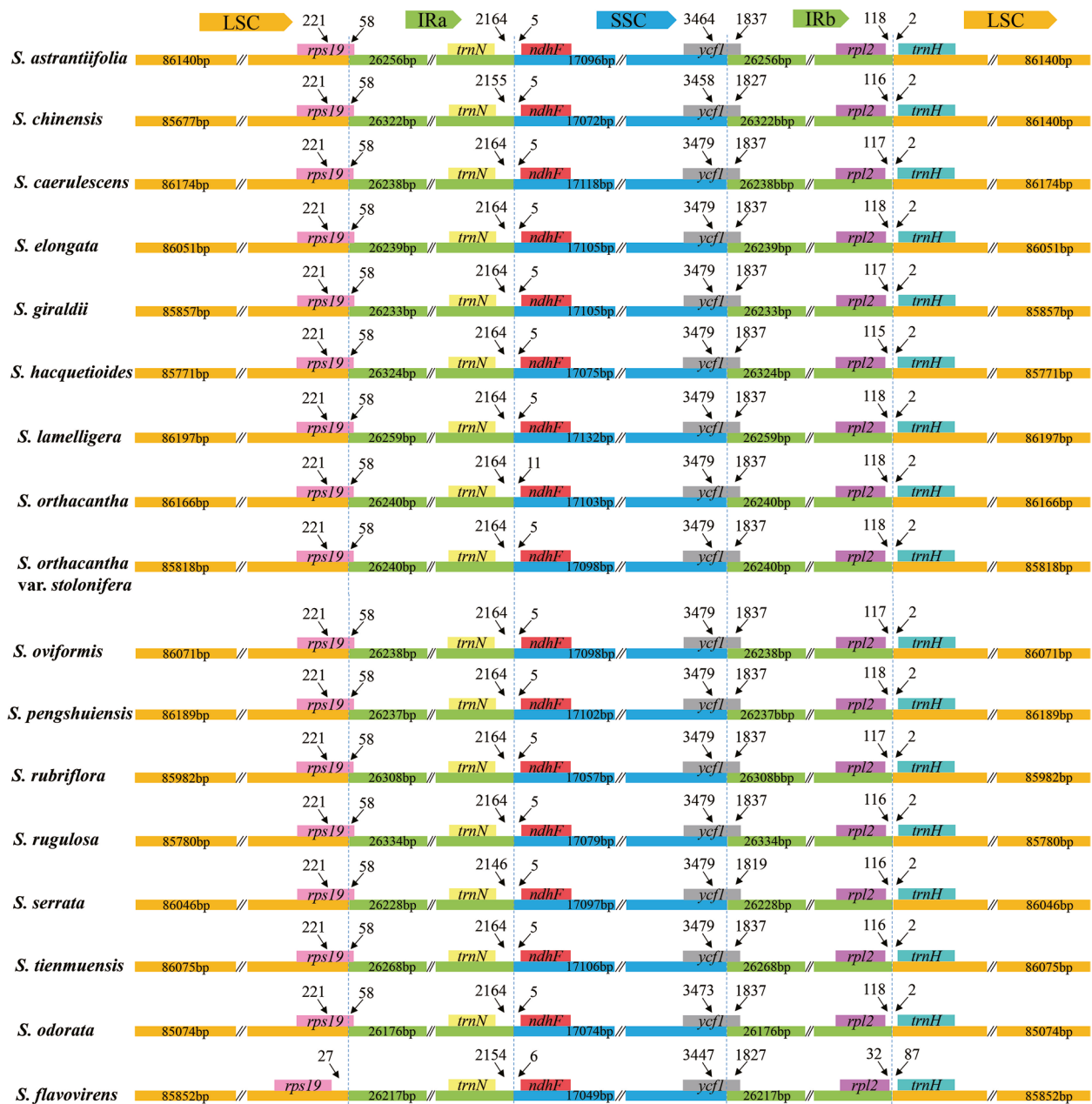


Fig. 4 Comparison of IR borders among the seventeen *Sanicula* plastomes. Different boxes for genes represent the gene position

87 bp away from the IRb region (Fig. 4). Thus, these findings indicated that seventeen *Sanicula* plastomes were highly conservative.

The Mauve result revealed that gene arrangement of the seventeen plastomes was highly conserved (Fig. 5). The result performed by the mVISTA program illustrated that the plastomes of seventeen taxa were also highly conserved (Fig. 6). In addition, the mVISTA

result found that the matrix contained 2472 variable sites (1.56%) and 1021 informative sites (0.65%) within the 15,8067 alignment positions. As a result, the whole plastomes sequence exhibited high similarity and no significant divergence was observed.



Fig. 5 Mauve alignment of seventeen *Sanicula* plastomes. Local collinear blocks within each alignment are represented by blocks of the same color connected with lines

Mutation hotspots in genus *Sanicula*

High nucleotide diversity (Pi) value means that certain regions have large variation and can be used to design the potential DNA barcodes. In this study, we calculated Pi values of four regions among the *Sanicula* plastomes. The results showed that there were 146 sites, including 107 sites in the LSC region, 19 sites in the SSC region and 20 sites in the IRs regions (Table S4). The findings also indicated that the Pi value of non-coding region was remarkably higher than that of the coding region (Fig. 7). Based on the sequence divergence, twelve mutation hotspot regions were selected as candidate DNA barcodes, containing seven protein coding genes (*atpF*, *cemA*, *accD*,

rpl22, *rbcL*, *matK* and *ycf1*) and five non-coding protein coding genes (*trnH-psbA*, *ycf4-cemA*, *rbcL-accD*, *trnE-trnT* and *trnG-trnR*). The seven protein coding genes with the highest sequence variation had obviously higher values of 0.00411 (0.79%), 0.00571 (1.16%), 0.00424 (0.7%), 0.00592 (1.02%), 0.00413 (0.85%), 0.00510 (0.73%) and 0.00691 (1.47%), respectively (Fig. 7A, Table S4). Among them, six genes were located in the LSC region and one gene (*ycf1* gene) was located at the SSC/IR junction. Five non-coding genes were considered as highly divergent hotspots with values of 0.03175 (7.07%), 0.01956 (0.99%), 0.02345 (2.81%), 0.03031 (1.99%) and

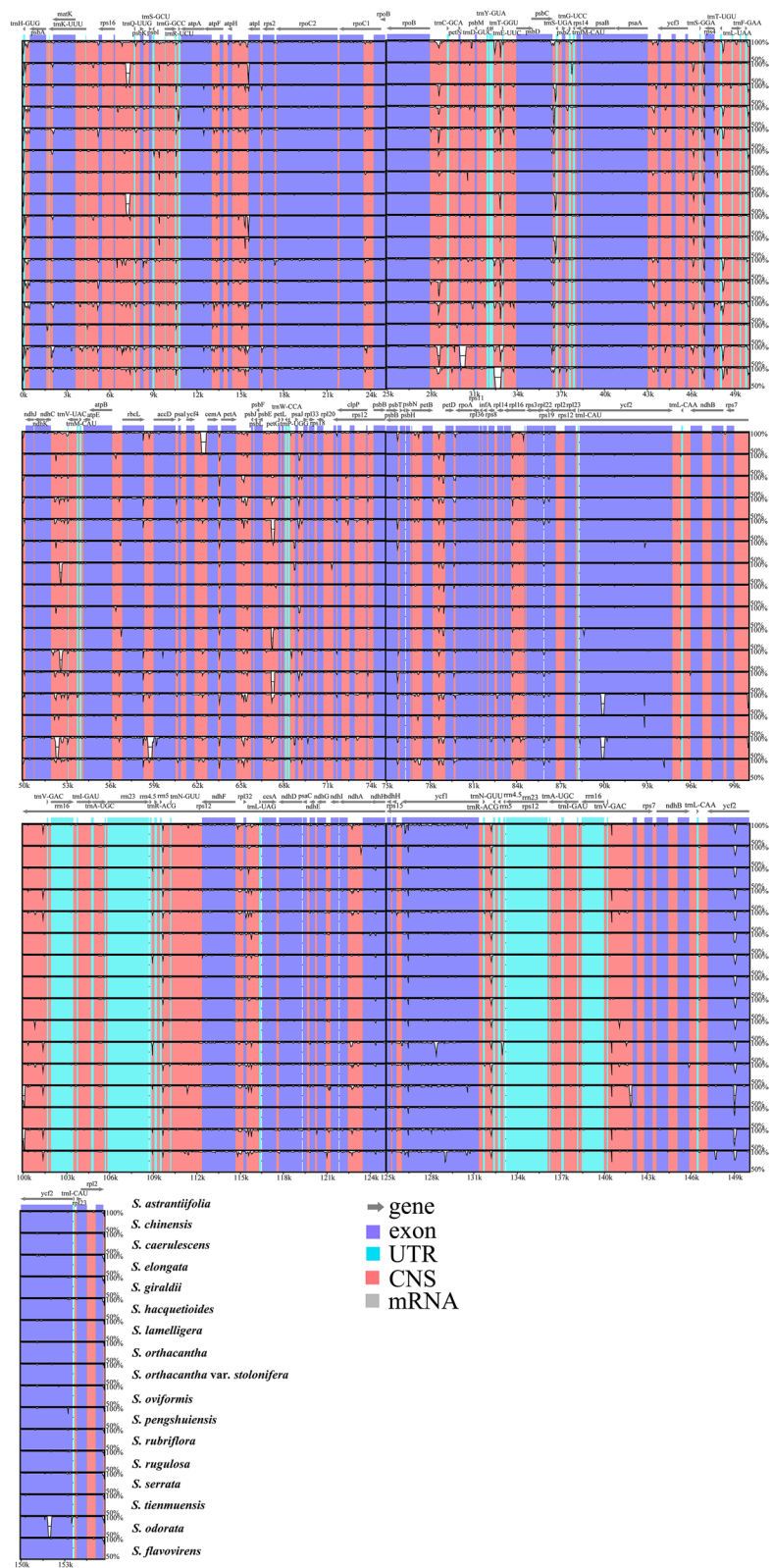


Fig. 6 mVISTA alignment for seventeen *Sanicula* plastomes with *S. astrantiifolia* as the reference

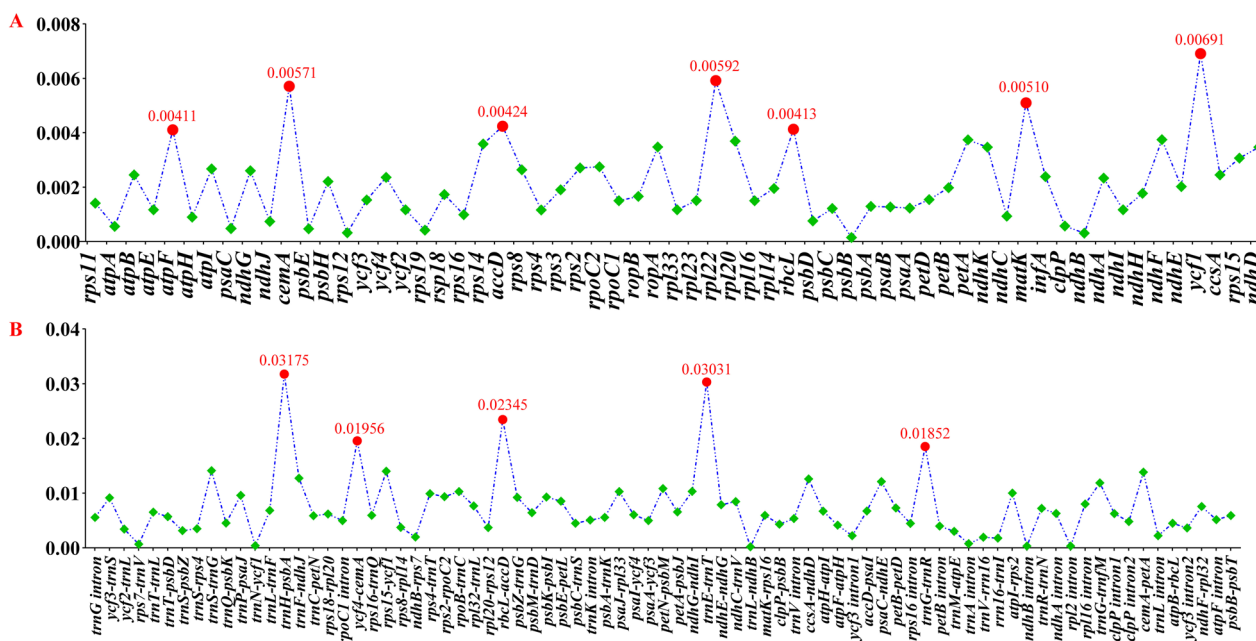


Fig. 7 The nucleotide diversity (Pi) values among seventeen *Sanicula* plastomes. **A** protein coding regions; **B** non-coding and intron regions

0.01852 (3%) (Fig. 7B, Table S4), all of which were located in the LSC region.

Phylogeny analyses for genus *Sanicula*

The 79 shared CDSs of 67 complete plastomes and 80 ITS sequences were performed to reconstruct the phylogeny of *Sanicula* (Table S5). Although the tree topologies of the plastome data and ITS sequences were incongruent, both strongly suggested that the monophyly of Saniculoideae and Apioideae. In addition, sister groups between Saniculoideae and Apioideae were robustly supported. The *Sanicula* species involved in this study clustered together and the *Eryngium* species also clustered together (Fig. 8).

In the CDS-based phylogenetic tree, the results of the maximum likelihood (ML) and Bayesian inference (BI) analyses yielded a well-resolved topologies and the topologies were highly identical as expected (Fig. 8A). It was clearly observed that the sections of *Sanicula* involved in the current study were not respectively recovered as monophyletic group. Instead, seventeen members of *Sanicula* were obviously divided into three clades. *S. rubriflora* F. Schmidt ex Maxim. clustered with *S. flavovirens* and *S. chinensis*, belonging to Clade I with robust support (PP = 1.00, BS = 100). Clade II included four species (*S. serrata*, *S. rugulosa*, *S. hacquetioides*, and *S. giraldii* H. Wolff.). The remaining ten species (*S. odorata*, *S. astrantiifolia*, *S. tienmuensis*, *S. orthacantha*, *S. oviformis* X.T. Liu & Z.Y. Liu., *S.*

caerulescens, *S. elongata* K.T. Fu., *S. pengshuiensis*, *S. lamelligera*, and *S. orthacantha* var. *stolonifera*) gathered together in Clade III, which could be divided into four subclades. It was noticed that Subclade A contained only one species: *S. odorata*. In Subclade B, *S. astrantiifolia* was sister to *S. tienmuensis* with weak supports (BS < 50, PP = 1.00). *S. orthacantha*, *S. oviformis* and *S. caerulescens* formed a strong Subclade C (PP = 1.00, BS = 100), of which *S. oviformis* and *S. caerulescens* clustered together and then resolved as sister to *S. orthacantha*. *S. elongata*, *S. pengshuiensis*, *S. lamelligera* and *S. orthacantha* var. *stolonifera* belonged to Subclade D, of which *S. elongata* clustered with *S. pengshuiensis* with more closely relationship (Fig. 8A).

The topologies of ITS tree obtained from ML and BI analyses were also highly identical (Fig. 8B). Although weakly supported in the ITS analyses, both analyses also suggested that the sections of *Sanicula* were not monophyletic group. The phylogenetic position of most species were consistent with the CDS-based phylogenetic trees, but there were a little conflicts. For example: (1) *S. elongata* clustered with *S. pengshuiensis* in the CDS-based phylogenetic trees (BS = 98, PP = 1.00), while the former formed a clade with *S. tienmuensis* in the ITS phylogenetic trees (BS = 94, PP = 1.00); (2) *S. rugulosa* was sister to *S. hacquetioides* in the CDS-based phylogenetic trees (BS = 100, PP = 1.00), whereas the former resolved sister to *S. giraldii* with depressed support (BS < 50, PP = 0.65) in the ITS phylogenetic trees; (3) *S. orthacantha* formed

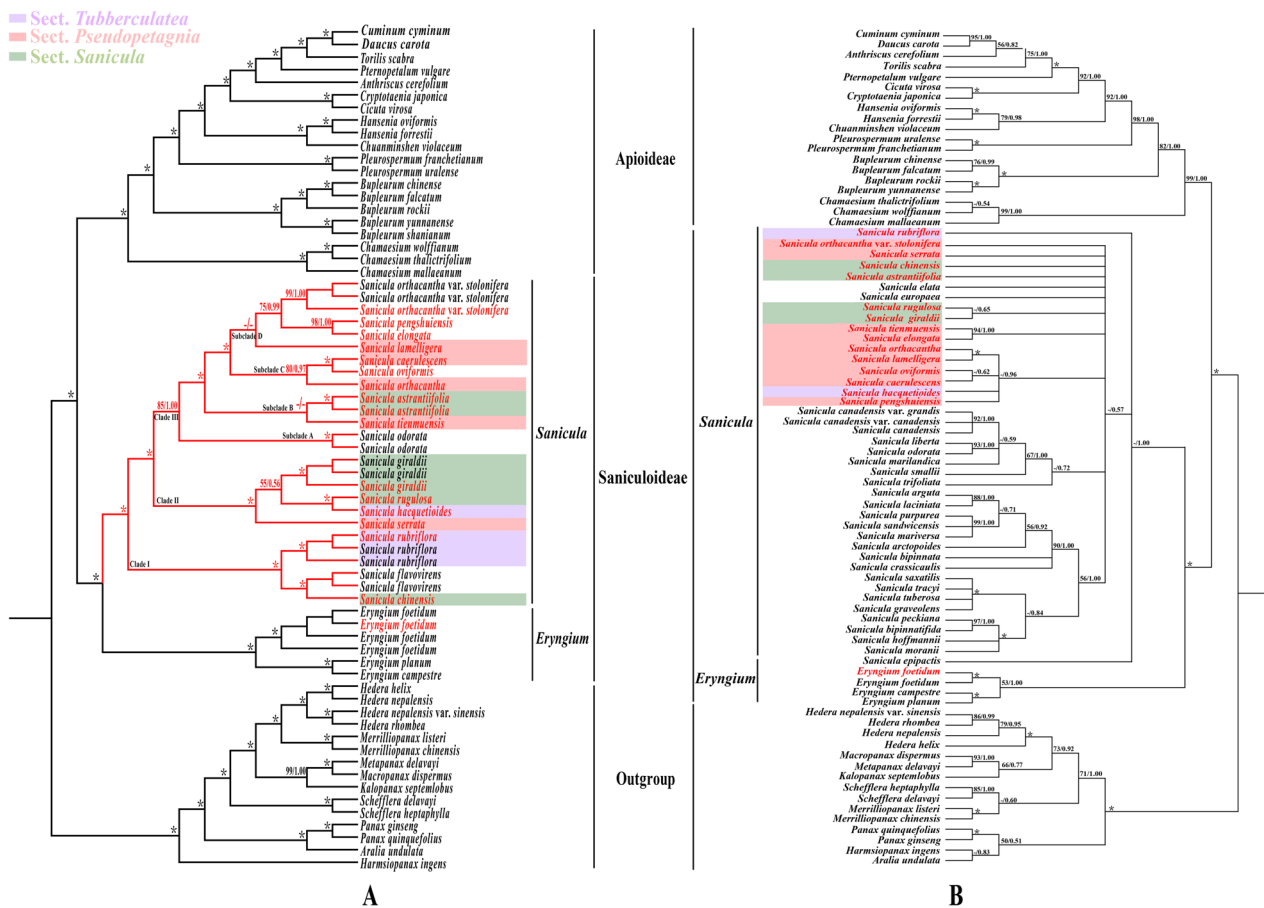


Fig. 8 Phylogenetic trees constructed by maximum likelihood (ML) and Bayesian inference (BI). The bootstrap values (BS) of ML and posterior probabilities (PP) of BI are listed at each node. (*) represents the node with PP = 1.00/ BS = 100. – means the values < 0.50/50. Red words indicates the newly sequenced species. **A** CDS tree; **B**: ITS tree

a clade with *S. oviformis* and *S. caerulescens* in the CDS-based phylogenetic trees (BS=80, PP=0.97), while *S. orthacantha* gathered together with *S. lamelligera* in the ITS phylogenetic trees with robust support (BS=100, PP=1.00); (4) *S. tienmuensis* clustered with *S. astrantiifolia* in the CDS-based phylogenetic trees (BS<50, PP<0.5), whereas the former was sister to *S. elongata* with strong support (BS=94, PP=1.00) in the ITS analyses (Fig. 8B).

Divergence time estimation

Based on four calibration points (one leaf macrofossil of Araliaceae and three calibration points constrained to a lognormal distribution), we estimated the divergence time of the genus *Sanicula*. Molecular dating results indicated that the family Apiaceae diverged from the family Araliaceae at 81.06 Ma (95% HPD: 53.97–87.24 Ma) in the Upper-Cretaceous. For the family Apiaceae, the divergence time between subfamily Apioideae and

subfamily Saniculoideae was estimated to be 64.19 Ma (95% HPD: 58.28–68.27) when was from the Upper-Cretaceous to the late Paleocene period. Subsequently, the subfamily Apioideae diverged at around 55.54 Ma (95% HPD: 47.77–63.39 Ma) in the early Eocene period. Within the subfamily Saniculoideae, the divergence time between the genus *Sanicula* and *Eryngium* occurred at 37.84 Ma (95% HPD: 20.33–52.21 Ma) in the late Eocene period. In addition, the diversification of *Sanicula* were happened at 18.38 Ma (95% HPD: 10.68–25.28 Ma) in the early Miocene period (Fig. 9).

Discussion

Comparative analysis of *Sanicula* plastomes

Comparative analyses of plastome can provide valuable insights into understanding patterns of molecular evolution [44, 45]. In this study, a comprehensive comparison of seventeen *Sanicula* plastomes were carried out. All *Sanicula* plastomes had a typical quadripartite structure,

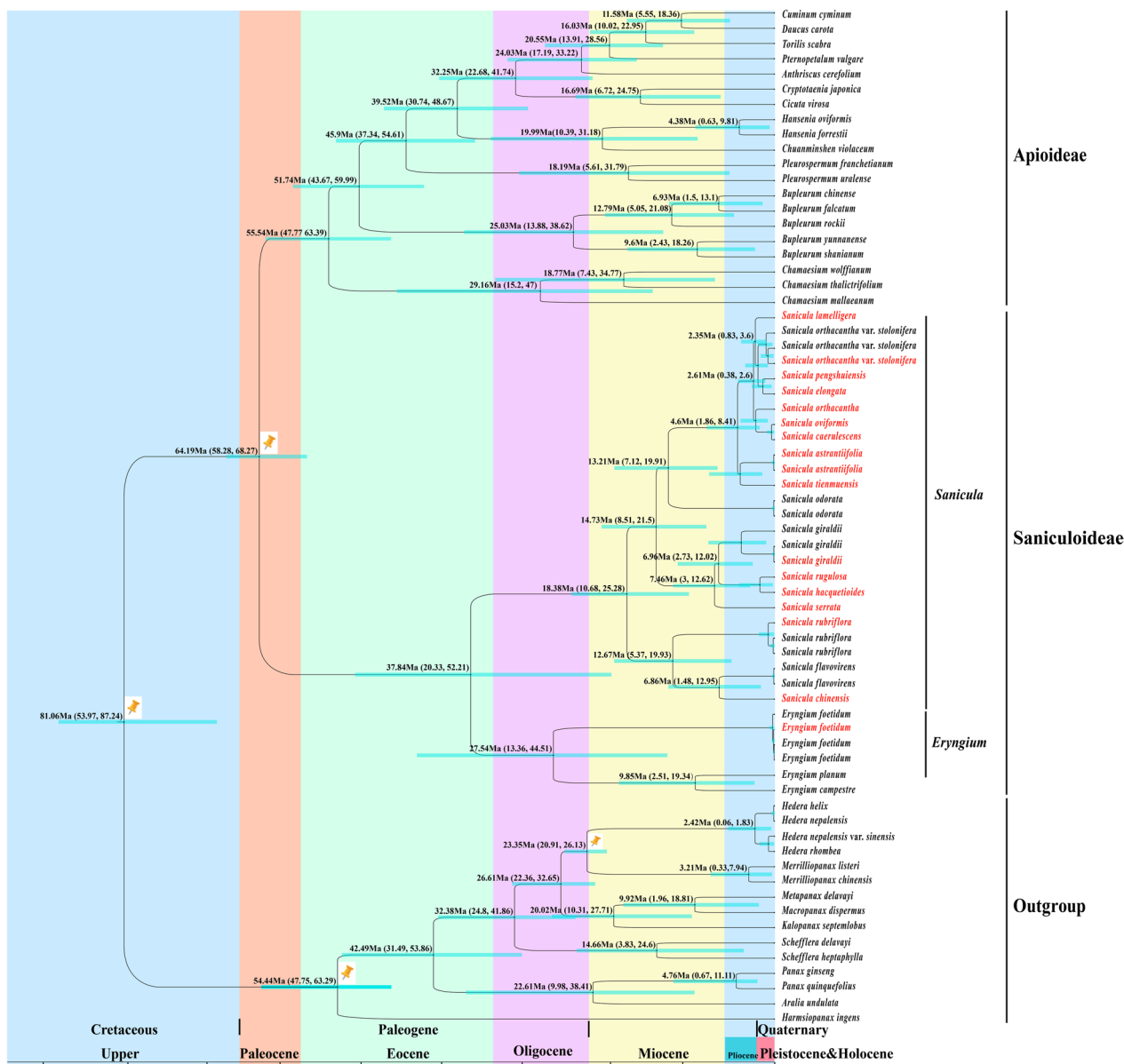


Fig. 9 The divergence time estimation based on 79 shared genes. The maximum credibility tree from the divergence times estimated with BEAST. The 95% highest posterior density (HPD) estimates for each well-supported clade are represented by bars. Orange nails indicate the calibration points for the molecular dating

consisting of one LSC region, one SSC region and a pair of IRs region (Fig. 1). Similarly, this structure was also detected in other plastomes of angiosperm plants [46–48]. Although gene loss is a common evolutionary event that frequently occurs in other genera of Apiaceae, such as the *ycf15* gene was lost in genus *Peucedanum* [31] and the *trnT-GGU* gene was lost in genus *Peucedanum* and *Kitagawia* [31, 35], it was not discovered in genus *Sanicula*. In addition, the gene arrangement and GC content of the seventeen plastomes displayed high similarity. These

findings demonstrated that the *Sanicula* plastomes were quite conserved. It is noteworthy that the GC content in the IR regions was relatively higher than that in the LSC and SSC regions (Table 1). The uneven distribution of GC content may be attributed to the higher GC content of the four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, and *rrn5*) in the IR regions, which is a common feature observed in other plants [49–51], as well as in Apiaceae [30–32].

IR contraction and expansion is a common phenomenon in the Apiaceae plastomes [20, 30–32, 35–40]. This

phenomenon was also observed in *Sanicula* plastomes. Our study displayed that the *rps19* gene had the length of 221 bp in the LSC region and 58 bp in the IRa region in sixteen plastomes, but in *S. flavovirens*, the entire *rps19* gene was located in the LSC region. In these seventeen plastomes (except *S. flavovirens*), the *rpl2* gene was 115–118 bp away from the IRb/LSC borders and the *trnH* gene was 2 bp away from the IRb/LSC borders, whereas in *S. flavovirens*, the *rpl2* gene was only 32 bp away from the IRb/LSC borders, and the *trnH* gene was 87 bp away from the IRb/LSC borders, which resulted in shorter plastome length of *S. flavovirens* than in the other sixteen *Sanicula* species. Therefore, we supported the hypothesis that the genome size variation was caused by IR contraction and expansion. Furthermore, we also examined the inverted repeats types described in the Apiaceae [20, 30–32, 35–40] and found that the situation observed in the genus *Sanicula* was also detected in the other genera of Apiaceae, such as in *Hansenia* Turcz., *Haplospheera* Hand.-Mazz., *Sinodielsia* H. Wolff [37].

Plastome evolution of *Sanicula*

Codons provide a link between nucleic acids and proteins, conveying genetic information [39]. The patterns of codon usage show basic characteristics of molecular evolution [52]. In species, codon usage is not random, as some synonymous codons are used more frequently than others, a phenomenon known as codon usage bias (CUB) [53]. CUB is considered to be a driver of gene evolution and it provides a great deal of information for understanding molecular evolution [54]. It exists in all species's genomes and is influenced by many factors, such as mutation pressure, natural selection, gene composition and gene length [55]. In the present study, all codons with RSCU > 1.00 were strongly biased towards A/U at the third codon position in seventeen *Sanicula* plastomes, which was consistent with the observation in most angiosperms [56], as well as in Apiaceae [30–32, 35–39]. Perhaps the high AT content in plastome is the major reason for the bias towards A/U for codons ending in A/U [39]. Leucine was the most abundant amino acid and was encoded by the highest number of codons, while cysteine was the rarest amino acid, as in most Apiaceae species [30–32].

Repeat sequences can cause gene insertion, deletion, substitution and duplication, which can lead to the generation of divergent regions in genome rearrangements. Therefore, these repeats can provide crucial information for phylogenetic and population studies [57]. Among the seventeen *Sanicula* plastomes, palindromic repeats were the majority of the four types of repeats, which have also been reported in the genus *Ligusticum* [30] and *Cnidium* [39].

SSRs are shorter tandem or microsatellite repeat sequences with repeat units of 1–6 bp, which are widely distributed in different regions of the plastome [58]. Due to a large amount of variation within the species, SSRs have been extensively used as valuable molecular markers for species authentication, population evolutionary analyses, phylogenetic relationship analyses, plant taxonomy, geography of species, genetic diversity and population structure studies [59]. Our research identified 979 SSRs in seventeen *Sanicula* plastomes. Mononucleotide were the most abundant SSRs, followed by dinucleotide, tetranucleotide, trinucleotide, hexanucleotide and pentanucleotide repeats. In addition, most of the SSRs contained A/T motifs, resulting in the AT richness of the seventeen *Sanicula* plastome. The similar pattern has also been observed in other Apiaceae plants [32]. In particular, the longest SSR was hexanucleotide (ACATAT/ATATGT) and it can serve as genetic markers in population genetics and phylogeography studies of *Sanicula* in the future.

Candidate DNA barcodes

With the rapid development of sequencing technologies and DNA barcoding, variable loci (e.g., *matK*, *rbcL* and *trnH-psbA*) have been recognized as universal DNA barcodes that were widely used for species identification, resource management, phylogenetic analyses and species evolution research in land plants [31, 36, 60, 61]. Although universal DNA barcoding have successfully distinguished many species, they have failed for some taxonomically notorious taxa and have encountered weak solutions [62]. Therefore, it is urgently needed to develop more effective variable loci as candidate DNA barcodes to solve the dilemma. Since plastome contain large-scale differentiation hotspots, thus it has the strong potential for screening DNA barcodes [63, 64]. In the current study, we investigated that the sequence variability in the non-coding region was higher than in the coding region, which was also detected in other genera of Apiaceae [31, 32]. Moreover, we identified five non-coding protein coding genes (*trnH-psbA*, *ycf4-cemA*, *rbcL-accD*, *trnE-trnT* and *trnG-trnR*) and seven protein coding genes (*atpF*, *cemA*, *accD*, *rpl22*, *rbcL*, *matK* and *ycf1*) that could be used as candidate DNA barcodes for *Sanicula* species identification. Of these twelve DNA barcodes, except for three universal DNA barcodes (*rbcL*, *matK* and *trnH-psbA*) and two promising DNA barcodes (*accD* and *ycf1*) in some plants [65], the remaining seven high variable regions (*atpF*, *cemA*, *rpl22*, *ycf4-cemA*, *rbcL-accD*, *trnE-trnT* and *trnG-trnR*) were different from those of other species [31, 66]. Therefore, they could be regarded as specific DNA barcodes to distinguish *Sanicula* species. These findings offered a valuable reference for further developing DNA barcodes for *Sanicula*.

Phylogeny inference and taxonomic implication of *Sanicula*

The plastome is uniparental inheritance, lacks recombination, and has highly variable loci; thus, it has been widely used in phylogenetic studies of angiosperms, especially at the low taxonomic levels of Apiaceae [20, 30–32, 35–40, 66, 67]. In the present study, the plastome data generated a well-resolved phylogenetic tree for *Sanicula*. The supports and resolutions of the plastome-based tree were significantly improved compared to those of the ITS-based tree constructed in the current and previous studies [5, 13, 15, 17].

The incongruences between ITS-based and plastome-based phylogenies are a common phenomenon in Apiaceae [31, 33, 35, 36, 67], and our results are no exception. These incongruences are mainly caused by hybridization/introgression, incomplete lineage sorting (ILS) and chloroplast capture events [20], as well as the maternal inheritance of plastid and biparental inheritance of ITS, which may also be an important factor [30].

Although topological incongruence existed between the plastome data and the ITS sequences, both successfully recognized all *Sanicula* species involved in the current study clustered together, which was consistent with several previous molecular phylogenetic analyses [5, 7, 12, 15, 17–19, 42]. Meanwhile, both topologies supported the sections within the genus were not being monophyletic group and the current classification may be inappropriate. Therefore, we considered that the current section-level classification system of *Sanicula* needs to be improved and revised. Although the current study did not provide a new taxonomic treatment of *Sanicula* taxa due to lack of insufficient sampling or the low supports and resolutions, our study laid the foundations and served as a framework for future taxonomic studies of *Sanicula*.

In addition, we clarified the taxonomic relationships of the four species. In the present study, both analyses indicated that *S. orthacantha* var. *stolonifera* was relatively far from *S. orthacantha*. Their notably different morphological characters of the two taxa further supported the phylogenetic results. For example, the rhizomes features (short, tuberlike, woody, oblique rootstock bearing elongated, fibrous roots, occasionally having fleshy stoloniferous) and narrow-linear, acute calyx teeth were observed in *S. orthacantha*, while slender, elongate and lignified nodes stoloniferous rhizomes and ovate calyx teeth were found in *S. orthacantha* var. *stolonifera* [13]. Thereby, *S. orthacantha* var. *stolonifera* should be regarded as an independent species rather than a variety of *S. orthacantha* based on the molecular and morphological evidence. A published study has treated *S. pengshuiensis* as

a synonymy of *S. lamelligera* [14]. Although the relationship between the two species was unsolved in ITS-based tree, the plastome-based tree robustly supported that *S. pengshuiensis* clustered with *S. elongata*. The morphological characteristics of them also supported the above phylogenetic results, such as inflorescence cymose branched, peduncles elongate was detected in *S. pengshuiensis*, while inflorescence 2–4-dichotomously or trichotomously branched, sometimes subcorymbose was found in *S. lamelligera* [13]. Hence, we suggested that *S. pengshuiensis* should be regarded as an independent species rather than a synonymy of *S. lamelligera*.

Estimation divergence time

Our analysis based on date estimation revealed that the divergence between the family Araliaceae and Apiaceae occurred during the Upper-Cretaceous period, with an approximate median age of 81.06 Ma (95% HPD: 53.97–87.24 Ma). This time was roughly consistent with the findings of Xie et al., who also used plastomes to estimate the divergence time of Apiales [40], but earlier than some studies (between 60 and 72 Ma [20, 68]). As for the family Apiaceae, the subfamilies Apioideae and Saniculoideae diverged at the beginning of the Paleogene, approximately 64.19 Ma (95% HPD: 58.28–68.27 Ma). The time was slightly earlier than other previous studies [20, 68, 69], but aligned closely with findings inferred from *trnD-trnT* + *rpl16* intron, with a deviation of 65.78 Ma (95% HPD: 58.21–74.31 Ma) [70]. The diversification of the subfamily Apioideae was estimated to have occurred during the early Eocene period 55.54 Ma. This time was earlier than that suggested by previous studies conducted by Wen et al. (43.45 Ma and 49.78 Ma) [20, 68]. On the other hand, in subfamily Saniculoideae, the genus *Sanicula* and *Eryngium* genera took place during the late Eocene period, precisely 37.84 Ma years ago (95% HPD: 20.33–52.21 Ma). This estimation was comparatively earlier than the previous records [20]. The potential discrepancy between our study and earlier previous researches may be caused by the differences in sampling and fossil calibrations used [40, 71]. Our results were considered more accurate and credible because the study contained a larger sample of *Sanicula* taxa and a broader range of calibration points, which were frequently employed in other evolutionary analyses [18, 40, 71, 72]. Therefore, we obtained a more precise divergence times.

Within subfamily Saniculoideae, the origin of genus *Sanicula* was estimated to have occurred during the late Eocene (37.84 Ma). Previous studies have reported that the initial uplift of the Qinghai-Tibetan Plateau (QTP) occurred at 45–35 Ma [71]. The QTP, which extends to the Hengduan Mountains (HDM) located on the south-eastern edge of the current QTP [73], encompasses the

most extensive range. Thus, the monolithic uplift of QTP led to the uplift of HDM, and some studies have suggested that the early uplift of the southern part of the HDM can be traced back to the period of late Paleocene or even the Eocene [74]. The uplift of the HDM resulted in the formation of a highly complex topography, giving rise to the creation of distinct and isolated “sky islands” [75, 76]. The “sky islands” have played a vital role in plant origin, speciation and evolution throughout the HDM and neighbouring regions, which has resulted in the emergence and diversification of various lineages, including *Plethodon ouachitae*, *Saussurea* and *Cupressus* L [77–80]. Therefore, we hypothesized that the uplift of the HDM may also have contributed to the origin of *Sanicula* species. The diversification of *Sanicula* occurred at 18.38 Ma, during the early Miocene. This period coincided with significant changes in East Asian monsoon intensity due to global temperature fluctuations [81], which led to substantial transformations in numerous biological communities in East Asia. This environmental shifts potentially facilitated the diversification of many species, such as *Primulina*, *Lepisorus* and *Begonia* [82–84]. Therefore, we assumed that the intensification of the East Asian monsoon might also acted as an indispensable role in promoting the diversification of *Sanicula*. Moreover, the uplift of HDM also likely influenced the diversification of this genus. The HDM experienced extensive uplift caused by the collision between the Indian plate and Eurasia, occurring after the Miocene and reaching its peak elevation before the Late Pliocene [85, 86]. This geological event triggered a series of topographical and climatic changes. The altered topography provided diverse habitats that facilitated species differentiation, as observed in the cases of *Caragana* and *Lilium* [87, 88]. Therefore, we supposed that the diversification of *Sanicula* species was closely related to the prevalence of the East Asian monsoon and the uplift of HDM events.

Conclusion

In this study, we successfully sequenced and assembled the complete plastomes of fifteen *Sanicula* species. Together with two previously reported plastomes, we performed a comprehensively comparative analyses. The results revealed that the genome structure, gene number, GC content, gene rearrangement and codon usage of seventeen *Sanicula* plastomes were highly conserved. Nevertheless, twelve highly variable regions were still selected as potentially strong DNA barcodes, which exhibited superior efficacy in species identification and phylogenetic relationship construction. Phylogenetic analyses indicated that the *Sanicula* species involved in this study clustered together. However, the existing section classification system was deemed non-natural. The

incongruences between ITS-based and plastome-based phylogenies were also observed in the genus. These discrepancies could largely be attributed to the events of hybridization/introgression, incomplete lineage sorting (ILS) and chloroplast capture, and the maternal inheritance of plastids, while the biparental inheritance of ITS. Molecular clock analysis indicated that the origin of genus *Sanicula* occurred during the late Eocene, which was significantly correlated with the uplift of the HDM. And the diversification of the genus occurred in the early Miocene, which was largely influenced by the prevalence of the East Asian monsoon and the uplift of the HDM. In conclusion, our study holds substantial significance and value for the investigation of the taxonomy, phylogeny, and evolution of the *Sanicula* genus.

Methods

Plant material and DNA extraction

The samples of fifteen *Sanicula* taxa were collected from the wild and then the fresh basal young leaves were immediately dried and stored with silica gel for further DNA extraction. The formal identification of these taxa was carried out by Professor Xingjin He (Sichuan University). Vouchers were stored in the herbarium of Sichuan University (Chengdu, China) under the deposition number listed in Table S6. Total DNA of newly collected samples was extracted from silica gel-dried fresh leaf tissues using the modified CTAB method [89]. We also newly sequenced fifteen *Sanicula* ITS sequences and submitted them to NCBI (accession numbers: OQ651137–OQ651152) (Table S5). To re-evaluate the taxonomic system of the genus more objectively, we also downloaded the plastomes and ITS sequences of this genus that were currently available on the NCBI.

Genome sequencing, assembly and annotation

The quality and quantity of DNA were tested using 1% agarose gel electrophoresis, and high-quality DNA was sequenced on an Illumina HiSeq2500 platform from Novogene (Beijing, China) according to the standard Illumina sequencing protocols [90]. Paired-end 150 reads were generated from libraries with an insert size of 300 bp. The raw data were filtered by software fastP v0.15.0 to obtain high-quality reads, with $-n$ 10 and $-q$ 15 [91]. For plastome assembly, we employed two strategies. First, we used NOVOPlasty v2.7.2 [92] to assemble complete plastomes, with the default parameters and *rbcL* sequence extracted from the plastome of *S. lamelligera* (MT561031) as the seed. To validate the accuracy of plastome assembly, GetOrganelle pipeline [93] was also used to assemble the plastomes, using the plastome sequence of *S. lamelligera* (GenBank accession: MT561031) as reference. And the same assembled results

were obtained by both methods. The assembled plastomes were initially annotated with the program DOGMA [94]. Then, the start and stop codons and intron positions were checked and manually corrected when necessary in Geneious v9.0.2 [95] and the plastomes of a given species recovered by different assembly strategies were also compared using Geneious v9.0.2 [95]. The circular plastome map was depicted by online program CHLOROPLOT (<https://irscope.shinyapps.io/Chloroplot/>) [96] and the newly sequenced and assembled seventeen complete plastomes were submitted into NCBI (accession numbers: OQ612639-OQ612643 and OQ626817-OQ626828) (Table S5).

Codon usage bias

The protein-coding genes (PCGs) were extracted from seventeen *Sanicula* plastomes for codon analysis. All overlapping genes were removed and the remaining 79 PCGs for each species. Since shorter genes may bias the codon usage estimation [97] and to avoid sampling bias [98], 79 PCGs with CDSs length shorter than 300 bp were excluded from this study, and 53 CDSs were finally screened. Relative synonymous codon usage (RSCU) is the value of the observed and expected frequency of a codon encoding a specific amino acid [99]. The RSCU value is divided into three ranges by the value 1.00. Among them, the RSCU value > 1.00 , $= 1.00$, and < 1.00 represents positive codon usage bias, no bias, and negative codon usage bias, respectively. Especially, codons with RSCU values greater than 1.6 and less than 0.6 are regarded as “over-represented” and “under-represented” codons, respectively [100]. The RSCU values were calculated using the CodonW v1.4.2 program [101] and the RSCU heatmap of the seventeen plastomes was emerged by GraphPad Prism 7 [102].

Repeat sequence and simple sequence repeats (SSRs) analysis

The four repetitive sequences: Forward (F), Reverse (R), Palindromic (P) and Complementary (C) repeats of the seventeen *Sanicula* plastomes were detected by the online REPuter program [103]. The parameters were set with maximum computed repeats $> 90\%$, minimal repeat size ≥ 30 bp, and a hamming distance = 3. All overlapping repeat sequences were removed. Simple sequence repeats (SSRs) were then checked using the MISA Perl script [104]. The minimum number of repeat units parameter was set to 10, 5, 4, 3, 3, and 3 for mono-, di-, tri-, tetra-, penta-, and hexanucleotides, respectively.

Comparative plastome of *Sanicula*

We compared the boundaries between inverted repeat and single-copy (IR/SC) among the seventeen plastomes

in Geneious v9.0.2 [95]. The junction regions of LSC/IRb/SSC/IRa were visualized by IRscope [105] after manually adjusted. Then, the DNA rearrangements among seventeen *Sanicula* plastomes were detected by using Mauve Alignment [106] implemented in Geneious v9.0.2 [95], with other parameters set as the default values. Furthermore, we also investigated the sequence divergence of *Sanicula* plastomes using mVISTA viewer [107], setting *S. astrantiifolia* as the reference.

Identification of high divergence hotspots

The protein coding regions, non-coding regions and intron regions of the seventeen plastomes were extracted in Geneious v9.0.2 [95] to identify mutation hotspot regions. We then aligned the sequences using MAFFT v7.221 [108] and the sequences were manually corrected with BioEdit software [109] and Geneious v9.0.2 [95]. And the alignment with less than 200 bp in length were discarded, mainly because the relatively short and insufficient sequence can yield the variation [110]. Finally, we evaluated nucleotide diversity (Pi) [111] using DnaSP version 5.0 software [112] to further investigate the molecular evolution of the *Sanicula* plastome.

Phylogenetic analysis

To illustrate the phylogenetic relationship of genus *Sanicula*, 67 plastomes and 80 ITS sequences were used to reconstruct the phylogenetic tree (Table S5). Among them, fifteen species of family Araliaceae were chosen as outgroup referred to the phylogenetic result of previous study [40]. The 80 ITS sequences were straightway aligned with MAFFT v7.221 [107] to gain the matrix. For plastome data, 79 commonly shared CDSs of 67 species were manually extracted in Geneious v9.0.2 [95] and aligned with MAFFT v7.221 [107]. Then the alignments were concatenated as a super matrix with PhyloSuite v1.2.2 [113]. Two matrixes were subjected to Maximum-Likelihood analysis (ML) and Bayesian Inference (BI). In detail, for the ML analyses, RAxML v8.2.8 [114] with GTRGAMMA model was performed to reconstruct the phylogenetic tree, and 1000 replicates was suggested to estimate the support value (BS) for each node according to the RAxML manual. Bayesian inference was performed by MrBayes v3.2.7 [115] and the best-fit nucleotide substitution model (GTR+I+G) for matrix of concatenated plastid protein-coding and (SYM+I+G) for ITS dataset were determine by Modeltest v3.7 [116] under the Akaike Information Criteria (AIC) [117]. The Markov chain Monte Carlo (MCMC) algorithm was run for 5,000,000 generations (sampling every 100 generations) with two runs and four chains (three heated chains and one cold chain). The running finished until the average standard deviation of split frequencies was below

0.01. An initial 25% of sampled trees were discarded, and the remainder were retained to generate the consensus tree and calculate posterior probabilities (PP). The results of ML and BI phylogenetic analyses were visualized in FigTree v1.4.2 [118].

Divergence time estimation

Bayesian relaxed clock analysis in program Bayesian Evolutionary Analysis Sampling Trees (BEAST v1.10.4) [119] was performed to estimate the divergence time of genus *Sanicula*. In the preliminary experiment, we analyzed random trees because the starting tree could not converge after two runs, but the ML phylogenetic tree was almost fully supported. Thus, the ML tree inferred from matrix of concatenated plastid protein-coding was used to fix topology. One leaf macrofossil of Araliaceae and three calibration points constrained to a lognormal distribution were used to constrain the phylogenetic tree. (i) the stem node of *Hedera* L. was constrained to a minimum age of 23.0 Ma based on leaf macrofossil of *Hedera* sp [120]. ; (ii) based on the fossilized leaves of Araliaceae, the stem node of *Harmiosipanax ingens* Philipson. was set as 49.28–72.9 Ma age [121]; (iii) according to the study of Wen et al., the root point was set to 68.88 Ma (95% HPD: 53.97–87.24 Ma) [20]; (iv) referring to the previous research of the evolutionary timescale, 54.17 Ma (95% HPD: 45.55–66.44) was implemented as a minimum age of subfamily Apioideae and Saniculoideae [20]. BEAUti was used to set criteria for analysis under the uncorrelated relaxed molecular clock model, with a Yule tree prior and the best-fit nucleotide substitution model (GTR + I + G) detected by Modeltest v3.7. Two independent replications of MCMC simulations were run for 2×10^9 generations with sampling every 2000 generations, with the first 20% trees being discarded as burn-in and the convergence of remaining runs was assessed through Tracer v1.7.1 [122] to examine the effective sample size (ESS) of all parameters not less than 200. The maximum clade credibility tree with median ages and 95% highest posterior density (HPD) intervals were produced with TreeAnnotator v2.1.2 [123] and the final result was exhibited in FigTree v1.4.2 [118].

Abbreviations

CTAB	Cetyl trimethylammonium bromide
PCR	Polymerase chain reaction
ITS	Internal transcribed spacer
CDS	Protein-coding sequences
bp	Base pair
BI	Bayesian inference
PP	Posterior probability
BS	Branch support
AIC	Akaike information criterion
ML	Maximum Likelihood

MCMC	Markov chain Monte Carlo
IR	Inverted repeat
LSC	Large single copy
SSC	Single copy
SSR	Simple sequence repeat
RSCU	Relative synonymous codon usage
Pi	Nucleotide diversity
rRNA	Ribosomal RNA
tRNA	Transfer RNA

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-04750-0>.

Additional file 1: Table S1. List of unique genes identified in plastomes of *Sanicula*.

Additional file 2: Table S2. Codon usage and relative synonymous codon usage (RSCU) values of 53 protein-coding genes of seventeen *Sanicula* plastomes.

Additional file 3: Table S3. Repeat sequence and simple sequence repeats in seventeen *Sanicula* plastomes.

Additional file 4: Table S4. Nucleotide diversity (Pi) of coding and non-coding regions.

Additional file 5: Table S5. The plastomes data included in phylogenetic analyses with GenBank accession.

Additional file 6: Table S6. The sample information of seventeen *Sanicula* in this study

Acknowledgements

We are grateful to Ting Ren, Jiaqing Lei and Huan-huan Qin for their valuable discussions. We thank Herbarium PE, KUN, SZ, NAS, CDBI for providing specimens consulting.

Authors' contributions

X-JH and S-DZ designed the work. B-NS, C-KL, A-QZ, Y-LX, and R-MT collected the materials; B-NS, C-KL, D-FX and H-C analyzed the data; B-NS wrote the manuscript; X-JH, S-DZ, C-KL, H-C and D-FX revised the manuscript. All authors gave final approval of the manuscript.

Funding

This research was funded by the National Natural Science Foundation of China (Grant No. 32070221, 32170209), and Survey on the Background Resources of Chengdu Area of Giant Panda National Park (Project No.: 510101202200376). The funders were not involved in the design of the research, collection, analysis and interpretation of data, and manuscript preparation.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the NCBI repository, <https://www.ncbi.nlm.nih.gov/>. Accession numbers: OQ612639-OQ612643, OQ626817- OQ626828 and OQ651137-OQ651152.

Declarations

Ethics approval and consent to participate

Collection of the fifteen samples completely complies with national and local legislation permission. Plant samples used in the study were not included in the list of national key protected plants and not collected from national park or natural reserve. According to national and local legislation, no specific permission was required for collecting these plants.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China.

²CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China.

Received: 8 August 2023 Accepted: 12 January 2024

Published online: 24 January 2024

References

- Pryer KM, Phillippe LR. A synopsis of the genus *Sanicula* (Apiaceae) in eastern Canada. *Can J Bot.* 1989;67(3):694–707.
- Pimenov MG, Leonov MV. The genera of the Umbelliferae: a nomenclator. Kew: Royal Botanic Gardens; 1993.
- Van Wyk B-E, Tilney PM, Magee AR. African Apiaceae: a synopsis of the Apiaceae/Umbelliferae of Sub-Saharan Africa and Madagascar/Ben-Erik Van Wyk, Patricia M Tilney & Anthony R Magee. Pretoria: Briza Academic Books; 2013.
- Koch WDJ. Generum tribiumque plantarum umbelliferarum nova dispositio. *Nova Acta Acad Caes Leop-Carol.* 1824;12:55–156.
- Vargas P, Baldwin BG, Constance L. A phylogenetic study of *Sanicula* sect. *Sanicoria* and *S.* sect. *Sandwicenses* (Apiaceae) based on nuclear rDNA and morphological data. *Syst Bot.* 1999;24(2):228.
- Linnaeus C. *Species Plantarum*. Stockholm: Laurentii Salvii; Oxford University Press; 1753.
- Calviño CI, Downie SR. Circumscription and phylogeny of apiaceae subfamily saniculoideae based on chloroplast DNA sequences. *Mol Phylogenet Evol.* 2007;44:175–91.
- Shan RH, Constance L. The genus *Sanicula* (Umbelliferae) in the old world and the new. *Univ Calif Press.* 1951;25:1–7.
- De Candolle AP. Umbelliferae. In: De Candolle AP, editor. *Prodromus systematis naturalis regni vegetabilis*, Vol. 4. Paris: Treüttel and Würzt; 1830. p. 55–220.
- De Candolle A. *Monographie des Campanulées*. Paris, Chez Mme. Veuve Desray, 1830.
- Drude Oscar. Umbelliferae. In: Engler, A. die natürlichen Pflanzenfamilien. Leipzig: Wilhelm Engelmann. 1898;37:63–128.
- Wolf H. Umbelliferae-Saniculoideae. In: Engler A, editor. *Das Pflanzenreich*. Volume IV. Leipzig & Berlin: Wilhelm Engelmann; 1913. pp. 1–305.
- Sheh ML, Phillippe LR. *Sanicula* L. In: Wu ZY, Raven PH, Hong DY, editors. *Flora of China*. Volume 14. Beijing Press. St. Louis: Missouri Botanical Garden Press; Science Press; 2005. pp. 19–24.
- Li HM, Song CF. Taxonomic studies on the genus *Sanicula* (Apiaceae) from China (I): the identity of *S. orthacantha* var. *Pumila* and *S. Pengshuiensis*. *Phytotaxa.* 2022;532(2):114–38.
- Valiejo-Roman CM, Terentjeva EI, Samigullin TH, Pimenov MG. Relationships among genera in Saniculoideae and selected Apioideae (Umbelliferae) inferred from nrITS sequences. *Taxon.* 2002;51(1):91–101.
- Yang C, Yao XY, Chen ZX, Downie SR, Wang QZ. The chloroplast genomes of *Sanicula* (Apiaceae): plastome structure, comparative analyses and phylogenetic relationships. *Nord J Bot.* 2022;2022(8):e03549.
- Calviño CI, Martínez SG, Downie SR. Morphology and biogeography of Apiaceae subfamily Saniculoideae as inferred by phylogenetic analysis of molecular data. *Am J Bot.* 2008;95(2):196–214.
- Vargas P, Baldwin BG, Constance L. Nuclear ribosomal DNA evidence for a western north American origin of hawaiian and south American species of *Sanicula* (Apiaceae). *PNAS.* 1998;95(1):235–40.
- Kadereit JW, Reppinger M, Schmalz N, Uhink CH, Wörz A. The phylogeny and biogeography of Apiaceae subf. Saniculoideae tribe Saniculoideae: from south to north and south again. *Taxon.* 2008;57(2):365–82.
- Wen J, Xie DF, Price M, Ren T, Deng YQ, Gui LJ, Guo XL, He XJ. Backbone phylogeny and evolution of Apioideae (Apiaceae): new insights from phylogenomic analyses of plastome data. *Mol Phylogenet Evol.* 2021;161: 107183.
- Standard of Chinese herbal medicines in Sichuan Province (Trial draft). Sichuan: Sichuan Health Bureau; 1979. p. 10.
- Li XS. Studies on the chemical constituents and biological activities of Feijiang Cao. Master's thesis. Hunan University of Traditional Chinese Medicine: Major of Chinese Medicine; 2011.
- Chen ZX. Taxonomic study on the genus *Sanicula* (Apiaceae) from China. Master's thesis. Huaqiao University: Major of Biology; 2019.
- Jansen RK, Ruhlman TA. Plastid genomes of seed plants. In: Bock R, Knoop V, editors. *Genomics of chloroplasts and mitochondria*. Dordrecht: Springer Netherlands; 2012. p 103–126.
- Ruhlman TA, Jansen RK. The plastid genomes of flowering plants. *Chloroplast Biotechnol.* 2014;1132:3–38.
- Mower JP, Vickrey TL. Structural diversity among plastid genomes of land plants. *Adv Bot Res.* 2018;85:263–92.
- Ravi V, Khurana JP, Tyagi AK, Khurana PJ. An update on chloroplast genomes. *Plant Syst Evo.* 2008;271:101–22.
- Palmer JD. Comparative organization of chloroplast genomes. *Annu Rev Genet.* 1985;19:325–354.
- Daniell H, Lin C, Yu M, Chang W. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 2016;17(1):134.
- Ren T, Li ZX, Xie DF, Gui LJ, Peng C, Wen J, He XJ. Plastomes of eight *Ligusticum* species: characterization, genome evolution, and phylogenetic relationships. *BMC Plant Biol.* 2020;20(1):1–14.
- Liu CK, Lei JQ, Jiang QP, Zhou SD, He XJ. The complete plastomes of seven *Peucedanum* plants: comparative and phylogenetic analyses for the *Peucedanum* Genus. *BMC Plant Biol.* 2022;22(1):101.
- Song BN, Liu CK, Xie DF, Xiao YL, Tian RM, Li ZX, Zhou SD, He XJ. Plastid phylogenomic analyses reveal the taxonomic position of *Peucedanum franchetii*. *Plants.* 2023;12(1):97.
- Li J, Cai J, Qin HH, Megan P, Zhang Z, Yu Y, Xie DF, Zhou SD, He XJ, Gao XF. Phylogeny, Age, and evolution of Tribe Liliae (Liliaceae) based on whole plastid genomes. *Front Plant Sci.* 2021;12:3158.
- Metzker ML. Sequencing technologies – the next generation. *Nat Rev Genet.* 2010;11(1):31–46.
- Lei JQ, Liu CK, Cai J, Price M, Zhou SD, He XJ. Evidence from Phylogenomics and morphology provide insights into the phylogeny, Plastome Evolution, and taxonomy of *Kitagawia*. *Plants.* 2022;11(23): 3275.
- Cai J, Qin HH, Lei JQ, Liu CK, He XJ, Zhou SD. The phylogeny of *Seseli* (Apiaceae, Apioideae): insights from molecular and morphological data. *BMC Plant Biol.* 2022;22(1):534.
- Gou W, Jia SB, Price M, Guo XL, Zhou SD, He XJ. Complete plastid genome sequencing of eight species from *Hansenia*, *Haplospheera* and *Sinodielsia* (Apiaceae): Comparative analyses and phylogenetic implications. *Plants.* 2020;9(11):1523.
- Tian RM, Aou XYM, Song BN, Li ZX, He XJ, Zhou SD. Plastid phylogenomic analyses reveal a cryptic species of *Ligusticopsis* (Apiaceae, Angiosperms). *Int J Mol Sci.* 2023;24:7419.
- Ren T, Aou X, Tian RM, Li ZB, Peng C, He XJ. Complete chloroplast genome of *Cnidium monnieri* (Apiaceae) and comparisons with other Tribe Selineae Species. *Diversity.* 2022;14(5):323.
- Xie DF, Xie C, Ren T, Song BN, Zhou SD, He XJ. Plastid phylogenomic insights into relationships, divergence, and evolution of Apiales. *Planta.* 2022;256(6):117.
- Chen ZX, Yao XY, Wang QZ. The complete chloroplast genome of *Sanicula chinensis*. *Mitochondrial DNA Part B.* 2019;4(1):734–5.
- Chen ZX, Yao XY, Downie SR, Wang QZ. Assembling and analysis of *Sanicula orthacantha* chloroplast genome. *Biodivers Sci.* 2019;27(4):366.
- Wang Z, Ren WC, Yan S, Zhang MQ, Liu YW, Ma W. Characterization of the complete chloroplast genome of *Sanicula rubriflora* F. Schmidt ex Maxim. *Mitochondrial DNA Part B.* 2021;67:1999–2000.
- Gitzendanner MA, Soltis PS, Yi TS, Li DZ, Soltis DE. Plastome phylogenetics: 30 years of inferences into plant evolution. In: *Adv Bot Res.* Academic Press; 2018. p. 293–313.
- Yao G, Jin JJ, Li HT, Yang JB, Mandala VS, Croyley M, Mostow R, Douglas AN, Chase WM, Christenhusz JM, Soltis DE, Soltis SP, Smith AS, Brockington SF, Moore MJ, Yi TS, Li DZ. Plastid phylogenomic insights into the evolution of *Caryophyllales*. *Mol Phylogenet Evol.* 2019;134:74–86.
- Tian XY, Guo J, Zhou XJ, Ma K, Ma YG, Shi TS, Shi YH. Comparative and evolutionary analyses on the complete plastomes of five *Kalanchoe* horticultural plants. *Front Plant Sci.* 2021;12:705874.
- Lee HO, Joh HJ, Kim K, Lee SC, Kim NH, Park JY, Park SH, Park MS, Kim S, Kwak M, Kim K, Lee WK, Yang TJ. Dynamic chloroplast genome

- rearrangement and DNA barcoding for three Apiaceae species known as the medicinal herb “Bang-Poong.” *Int J Mol Sci.* 2019;20(9):2196.
48. Wang Y, Zhang CF, Odago WO, Jiang H, Yang JX, Hu GW, Wang QF. Evolution of 101 Apocynaceae plastomes and phylogenetic implications. *Mol Phylogenet Evol.* 2023;180:107688.
 49. Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kologanova TV, Skyabin KG. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J Mol Evol.* 2008;66:555–64.
 50. Jiang DZ, Cai XD, Gong M, Xia MQ, Xing HT, Dong SS, Tian SM, Li JL, Lin JY, Liu YQ, Li HL. Complete chloroplast genomes provide insights into evolution and phylogeny of *Zingiber* (Zingiberaceae). *BMC Genom.* 2023;24(1):1–18.
 51. Niu ZT, Xue QY, Wang H, Xie XZ, Zhu SY, Liu W, Ding XY. Mutational biases and GC-biased gene conversion affect GC content in the plastomes of *Dendrobium* Genus. *Int J Mol Sci.* 2017;18(11):2307.
 52. Yan Q, Philmus B, Hesse C, Kohlen M, Chang JH, Loper JE. The rare codon AGA is involved in regulation of pyoluteorin biosynthesis in *Pseudomonas protegens* Pf-5. *Front Microbiol.* 2016;7:497.
 53. Sharp PM, Li WH. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J Mol Evol.* 1986;24:28–38.
 54. Mitreva M, Wendl MC, Martin J, Wylie T, Yin Y, Larson A, Parkinson J, Waterston RH, McCarter JP. Codon usage patterns in Nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Bio.* 2006;7(8):1–19.
 55. Salim HM, Cavalcanti AR. Factors influencing codon usage bias in genomes. *J Braz Chem Soc.* 2008;19:257–62.
 56. Maheswari P, Kunhikannan C, Yasodha R. Chloroplast genome analysis of angiosperms and phylogenetic relationships among Lamiaceae members with particular reference to teak (*Tectona grandis* Lf.). *J Biosci.* 2021;46:1–9.
 57. Wang X, Xu YF, Fan HY, Cui N, Meng XN, He JJ, Ran NN, Yu Y. Research progress of plant nucleotide-binding leucine-rich repeat protein. *Horticulturae.* 2023;9(1):122.
 58. Chen C, Zhou P, Choi YA, Huang S, Gmitter FG. Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet.* 2006;112:1248–57.
 59. Echt CS, DeVerno LL, Anzidei M, Vendramin GG. Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Mol Ecol.* 1998;7(3):307–16.
 60. Gregory TR. DNA barcoding does not compete with taxonomy. *Nature.* 2005;434(7037):1067–1067.
 61. Li X, Li YF, Zang MY, Li MZ, Fang YM. Complete chloroplast genome sequence and phylogenetic analysis of *Quercus acutissima*. *Int J Mol Sci.* 2018;19(8):2443.
 62. Cao Q, Xu LH, Wang JL, Zhang FQ, Chen SL. Molecular phylogeny of subtribe Swertiinae. *BBR.* 2021;41(03):408–18.
 63. Chen XL, Zhou JG, Cui YX, Wang Y, Duan BZ, Yao H. Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode. *Front Pharmacol.* 2018;9:695.
 64. Li XW, Yang Y, Henry RJ, Rossetto M, Wang YT, Chen SL. Plant DNA barcoding: from gene to genome. *Biol Rev.* 2015;90(1):157–66.
 65. Duchene D, Bromham L. Rates of molecular evolution and diversification in plants: chloroplast substitution rates correlate with species richness in the Proteaceae. *BMC Evol Biol.* 2013;13(1):1–11.
 66. Li ZX, Guo XL, Price M, Zhou SD, He XJ. Phylogenetic position of *Ligusticopsis* (Apiaceae, Apioidaeae): evidence from molecular data and carpological characters. *AoB Plants.* 2022;14(2):plac008.
 67. Qin HH, Cai J, Liu CK, Zhou RX, Price M, Zhou SD, He XJ. The plastid genome of twenty-two species from *Ferula*, *Talassia*, and *Soranthus*: comparative analysis, phylogenetic implications, and adaptive evolution. *BMC Plant Biol.* 2023;23(1):1–18.
 68. Wen J, Yu Y, Xie DF, Peng C, Liu Q, Zhou SD, He XJ. A transcriptome based study on the phylogeny and evolution of the taxonomically controversial subfamily Apioidaeae (Apiaceae). *Ann Bot.* 2020;125:937–53.
 69. Calviño CI, Teruel FE, Downie SR. The role of the Southern Hemisphere in the evolutionary history of Apiaceae, a mostly north temperate plant family. *J Biogeogr.* 2016;43:398–409.
 70. Nicolas AN, Plunkett GM. Diversification times and biogeographic patterns in Apiales. *Bot Rev.* 2014;80:30–58.
 71. Barba-Montoya J, Reis MD, Schneider H, Donoghue PCJ, Yang Z. Constraining uncertainty in the timescale of angiosperm evolution and the veracity of a cretaceous Terrestrial Revolution. *New Phytol.* 2018;218(2):819–34.
 72. Valcárcel V, Wen J. Chloroplast phylogenomic data support Eocene amphipacific early radiation for the Asian Palmate core Araliaceae. *J Syst Evol.* 2019;57(6):547–60.
 73. Zhang Y, Li B, Zheng D. Datasets of the boundary and area of the Tibetan Plateau. *Acta Geogr Sin.* 2014;69(5):164–8.
 74. Wu SD, Wang Y, Wang ZF, Shrestha N, Liu JQ. Species divergence with gene flow and hybrid speciation on the Qinghai–Tibet Plateau. *New Phytol.* 2022;234(2):392–404.
 75. Lopez-Pujol J, Zhang FM, Sun HQ, Ying TS, Ge S. Centres of plant endemism in China: places for survival or for speciation? *J Biogeogr.* 2011;38:1267–80.
 76. Sklenář P, Hedberg I, Cleef AM. Island biogeography of tropical alpine floras. *J Biogeogr.* 2014;41(2):287–97.
 77. Shepard DB, Burbrink FT. Lineage diversification and historical demography of a sky island salamander, *Plethodon ouachitae*, from the Interior Highlands. *Mol Ecol.* 2008;17(24):5315–35.
 78. Wang YJ, Susanna A, Von Raab-Straube E, Milne R, Liu JQ. Island-like radiation of *Saussurea* (Asteraceae: Cardueae) triggered by uplifts of the Qinghai–Tibetan Plateau. *Biol J Linn Soc.* 2009;97(4):893–903.
 79. Xu TT, Abbott RJ, Milne RI, Mao KS, Du FK, Wu GL, Ciren ZX, Miehle G, Liu JQ. Phylogeography and allopatric divergence of cypress species (*Cupressus* L.) in the Qinghai-Tibetan Plateau and adjacent regions. *BMC Evol Biol.* 2010;10:1–10.
 80. Qiu YX, Fu CX, Comes HP. Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of quaternary climate and environmental change in the world’s most diverse temperate flora. *Mol Phylogenet Evol.* 2011;59:225–44.
 81. Li SF, Valdes PJ, Farnsworth A, Davies-Barnard T, Su T, Lunt DJ, Spicer RA, Liu J, Deng WYD, Huang J, Tang H, Ridgwell A, Chen LL, Zhou ZK. Orographic evolution of northern Tibet shaped vegetation and plant diversity in eastern Asia. *Sci Adv.* 2021;7(5):eabc7741.
 82. Kong H, Condamine FL, Harris AJ, Chen J, Pan B, Möller M, Hoang VS, Kang M. Both temperature fluctuations and east Asian monsoons have driven plant diversification in the karst ecosystems from southern China. *Mol Ecol.* 2017;26(22):6414–29.
 83. Wang L, Schneider H, Zhang XC, Xiang QP. The rise of the Himalaya enforced the diversification of SE Asian ferns by altering the monsoon regimes. *BMC Plant Biol.* 2012;12:1–9.
 84. Chung KF, Leong WC, Rubite RR, Repin R, Kiew R, Liu Y, Peng CI. Phylogenetic analyses of *Begonia* sect. *Coelocentrum* and allied limestone species of China shed light on the evolution of sino-vietnamese karst flora. *Bot Stud.* 2014;55:1–15.
 85. Zhao YJ, Yin GS, Pan YZ, Tian B, Gong X. Climatic refugia and geographical isolation contribute to the speciation and genetic divergence in Himalayan-Hengduan tree peonies (*Paeonia delavayi* and *Paeonia ludlowii*). *Front Genet.* 2021;11:595334.
 86. Dupont-Nivet G, Lippert PC, Van Hinsbergen DJJ, Meijers MJM, Kapp P. Palaeolatitude and age of the indo–Asia collision: palaeomagnetic constraints. *Geophys J Int.* 2010;182:1189–98.
 87. Zhang ML, Fritsch PW. Evolutionary response of *Caragana* (Fabaceae) to Qinghai–Tibetan Plateau uplift and Asian interior aridification. *Plant Syst Evol.* 2010;288:191–9.
 88. Gao YD, Harris AJ, Zhou SD, He XJ. Evolutionary events in *Lilium* (including *Nomocharis*, Liliaceae) are temporally correlated with orogenies of the QT plateau and the Hengduan Mountains. *Mol Phylogenet Evol.* 2013;68(3):443–60.
 89. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bull.* 1987;19:11–5.
 90. Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26:1135–45.
 91. Chen SF, Zhou YQ, Chen YR, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–890.
 92. Dierckxens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *NAR.* 2017;45(4):e18–e18.

93. 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):1–31.
94. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 2004;20:3252–5.
95. 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.
96. Zheng S, Poczaï P, Hyvönen J, Tang J, Amiryousefi A. Chloroplot: an online program for the versatile plotting of organelle genomes. *Front Genet.* 2020;11: 576124.
97. Kessler MD, Dean MD. Effective population size does not predict codon usage bias in mammals. *Ecol Evol.* 2014;4:3887–900.
98. Yang Y, Zhu J, Feng L, Tao Z, Bai G, Yang J, Zhao G. Plastid genome comparative and phylogenetic analyses of the key genera in *Fagaceae*: highlighting the effect of codon composition bias in phylogenetic inference. *Front Plant Sci.* 2018;9:82.
99. Wu HG, Bao ZY, Mou CX, Chen ZH, Zhao JW. Comprehensive analysis of codon usage on porcine Astrovirus. *Viruses.* 2020;12:991.
100. Wong EH, Smith DK, Rabadan R, Peiris M, Poon LL. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. *BMC Evol Biol.* 2010;10: 253.
101. Peden JF. Analysis of codon usage. PhD thesis. Nottingham University: Department of Genetics; 1999.
102. Mitteer DR, Greer BD, Fisher WW, Cohrs VL. Teaching behavior technicians to create publication-quality, single-case design graphs in graphpad prism 7. *J Appl Behav Anal.* 2018;51(4):998–1010.
103. Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. *NAR.* 2001;29(22):4633–42.
104. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). *Theor Appl Genet.* 2003;106:411–22.
105. Amiryousefi A, Hyvonen J, Poczaï P. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics.* 2018;34:3030–1.
106. Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 2004;14(7):1394–403.
107. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* 2004;32(suppl2):W273–279.
108. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
109. Hall AT. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic Acids Symp Ser.* 1999; p. 95–98.
110. Hong SY, Cheon KS, Yoo KO, Lee HO, Cho KS, Suh JT, Kim SJ, Nam JH, Sohn GB, Kim YH. Complete chloroplast genome sequences and comparative analysis of *Chenopodium quinoa* and *C. album*. *Front Plant Sci.* 2017;8:1696.
111. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci.* 1979;76:5269–73.
112. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25(11):1451–2.
113. 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.
114. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3.
115. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard M, Huelsenbeck JP. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–42.
116. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics.* 1998;14(9):817–8.
117. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst Biol.* 2004;53(5):793–808.
118. Rambaut A, Drummond A. FigTree, version 1.4.2. 2015. <http://tree.bio.ed.ac.uk/software/figtree/>. Accessed 4 Jun 2023.
119. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 2018;4:vey016.
120. Kong WS. Vegetational history of the Korean Peninsula. *Global Eco Biogeogr.* 2000;9:391–402.
121. Manchester SR. Fruits and seeds of the middle Eocene nut beds Flora, Clarno formation, Oregon. *Palaeontogr Am.* 1994;58:38–9.
122. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 2018;67(5):901–4.
123. Rambaut A, Drummond AJ. TreeAnnotator v2.1.2. Edinburgh: University of Edinburgh, Institute of Evolutionary Biology; 2014.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.