BMC Genomics

RESEARCH

Open Access

Comparative analysis of complete chloroplast genomes of *Synotis* species (Asteraceae, Senecioneae) for identification and phylogenetic analysis

Xiaofeng Liu[1,2], Junjia Luo[2], Hui Chen[2], Tingyu Li[2], Tianmeng Qu[2], Ming Tang[3*] and Zhixi Fu[1,2,4*]

Abstract

Background The *Synotis* (C. B. Clarke) C. Jeffrey & Y. L. Chen is an ecologically important genus of the tribe Senecioneae, family Asteraceae. Because most species of the genus bear similar morphology, traditional morphological identification methods are very difficult to discriminate them. Therefore, it is essential to develop a reliable and effective identification method for *Synotis* species. In this study, the complete chloroplast (cp.) genomes of four *Synotis* species, *S. cavaleriei* (H.Lév.) C. Jeffrey & Y.L. Chen, *S. duclouxii* (Dunn) C. Jeffrey & Y.L. Chen, *S. nagensium* (C.B. Clarke) C. Jeffrey & Y.L. Chen and *S. erythropappa* (Bureau & Franch.) C. Jeffrey & Y. L. Chen had been sequenced using next-generation sequencing technology and reported here.

Results These four cp. genomes exhibited a typical quadripartite structure and contained the large single-copy regions (LSC, 83,288 to 83,399 bp), the small single-copy regions (SSC, 18,262 to 18,287 bp), and the inverted repeat regions (IR, 24,837 to 24,842 bp). Each of the four cp. genomes encoded 134 genes, including 87 protein-coding genes, 37 tRNA genes, 8 rRNA genes, and 2 pseudogenes (*ycf1* and *rps19*). The highly variable regions (*trnC-GCA-petN*, *ccsA-psaC*, *trnE-UUC-rpoB*, *ycf1*, *ccsA* and *petN*) may be used as potential molecular barcodes. The complete cp. genomes sequence of *Synotis* could be used as the potentially effective super-barcode to accurately identify *Synotis* species. Phylogenetic analysis demonstrated that the four *Synotis* species were clustered into a monophyletic group, and they were closed to the *Senecio*, *Crassocephalum* and *Dendrosenecio* in tribe Senecioneae.

Conclusions This study will be useful for further species identification, evolution, genetic diversity and phylogenetic studies within this genus *Synotis* and the tribe Senecioneae.

Keywords *Synotis*, Chloroplast genome, Super-barcode, Comparative analysis, Phylogenetic analysis

*Correspondence:
Ming Tang
tangming@jxau.edu.cn
Zhixi Fu
fuzx2017@sicnu.edu.cn
[1]Key Laboratory of Land Resources Evaluation and Monitoring in Southwest (Sichuan Normal University), Ministry of Education, Chengdu 610066, China
[2]College of Life Sciences, Sichuan Normal University, Chengdu 610101, China
[3]Jiangxi Provincial Key Laboratory for Bamboo Germplasm Resources and Utilization, College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China
[4]Sustainable Development Research Center of Resources and Environment of Western Sichuan, Sichuan Normal University, Chengdu 610101, China

Background

The genus *Synotis* (C.B. Clarke) C. Jeffrey & Y.L. Chen belongs to tribe Senecioneae of the family Asteraceae. It typically comprises subshrubs and perennial herbs. The leaves of *Synotis* are simple, either petiolate or sessile, with broadly ovate-cordate to narrowly oblong-lanceolate plates. Its capitula are heterogamous and radiate, always arranged in terminal or axillary [1]. It consists of approximately 60 species, mainly distributed in Bhutan, China, India, Myanmar, Nepal, Thailand, and Vietnam [1–13]. China, the diversity center of *Synotis*, harbors approximately 50 species, and approximately 30 species are endemic [6–9, 11–13]. The genus was separated from *Senecio* L. mainly because of its conspicuously caudate anthers [14]. Tang identified five series within *Synotis*, including ser. *Synotis*, ser. *Erectae* (C. B. Clarke) C. Jeffrey & Y. L. Chen, ser. *Oliganthae* (J. F. Jeffrey) C. Jeffrey & Y. L. Chen, ser. *Fulvipapposae* C. Jeffrey & Y. L. Chen, and ser. *Hieraciifoliae* M. Tang & Q. E. Yang [6]. Most species in this genus exhibit relatively similar morphological variations, making it challenging to distinguish them using traditional identification methods.

During our fieldwork in southwestern China, the materials of four species of *Synotis*, including *S. cavaleriei* (ser. *Hieraciifoliae*), *S. duclouxii* (ser. *Hieraciifoliae*), *S. nagensium* (ser. *Erectae*) and *S. erythropappa* (ser. *Oliganthae*) were collected. Among these species, *S. nagensium* is mainly distributed in China (Chongqing, Gansu, Guangdong, Guangxi, Guizhou, Hubei, Hunan, Sichuan, Xizang, Yunnan provinces), as well as in India, Nepal, Bhutan, Myanmar, Thailand and Vietnam [1, 11]. The other three species are endemic to China [1], of which *S. cavaleriei* is mainly distributed in Guizhou, Sichuan, and Yunnan provinces [1, 6, 8], *S. duclouxii* is found mainly in Sichuan and Yunnan provinces [1, 8], and *S. erythropappa* mainly occurs in Guizhou, Hubei, Sichuan, Xizang, and Yunnan provinces [1]. These four species often grow in mixed forests, woods, thickets and meadows [1]. They play a crucial role in ecological restoration and soil and water conservation efforts. The species of *S. nagensium* has disciform capitula and paniculoid thyrses, whereas *S. erythropappa* has broadly pyramidal compound thyrses, which two are easily distinguished from each other. However, the other two species bear the same habit, leaf and inflorescence shape, making them difficult to distinguish by traditional identification (Fig. 1).

The chloroplast (cp.) is a vital organelle in green plants. It converts light energy to chemical energy through photosynthesis [15]. The chloroplast contains its own genomes [16]. In general, a typical cp. genome of the family Asteraceae contains a quadripartite architecture, with two inverted repeat regions (IRs) that separate a large single-copy region (LSC) and a small single-copy region (SSC) [17–19]. The cp. genomes generally contain 100–120 genes with lengths of 120–160 kb [15, 20, 21]. Compared with the partial cp. genes and nuclear genes, the complete cp. genomes exhibit unique characteristics, such as conserved genetics, slow evolution, maternal inheritance, and numerous mutation site information [18, 19, 22]. Consequently, the complete cp. genomes have become an efficient tool for phylogenetic studies within the family Asteraceae [23, 24] and serve as a super-barcode for distinguishing species [25–29]. The advent of high-throughput sequencing has facilitated the reporting of several cp. genomes within the Asteraceae family, including *Artemisia* L. [13, 18, 22, 30], *Aster* L. [31–36], *Cavea* W. W. Smith and J. Small [24], *Conyza* L. [26, 37], *Chrysanthemum* L. [38], *Crepidiastrum* Nakai [39], *Dolomiaea* DC. [40], *Erigeron* L. [41], *Gynura* Cass. [42], *Heteroplexis* C.C.Chang [43], *Ligularia* Cass. [25], *Nouelia* Franch. [44], *Saussurea* DC. [17, 45], *Sinosenecio* B. Nordenstam [19], *Xanthium* L. [46]. These studies have contributed to enhancing our understanding of the cp. genome characteristics and evolutionary relationships within the Asteraceae family.

With the rapid development of molecular technology, external transcribed spacers (ETS) [47–50], internal transcribed spacers (ITS) [47–51] and cp. fragments, such as *matK-trnK-rpS16*, *rpS16-trnQ-psbK* have been employed to determine the molecular phylogeny of *Synotis* [6]. However, the complete cp. genomes of genus *Synotis* have not yet been reported. Compared to nuclear markers and partial cp. genes, the complete cp. genomes possess highly conserved DNA sequences and many genetic information. Therefore, analyzing the complete cp. genomes may be an effective approach to solve the problem of the evolutionary relationships and taxonomic identification of species [18, 19, 22]. Meanwhile, with advances in next-generation sequencing technology, it has recently become quicker, cheaper and easier to achieve the complete plastomes than before [52–54]. Therefore, it is crucial to obtain the complete cp. genomes of *Synotis* for reconstructing phylogenetic relationships and species identification.

In this study, the complete cp. genomes of *Synotis cavaleriei*, *S. duclouxii*, *S. nagensium*, and *S. erythropappa* were sequenced using Illumina technology, and their gene features were characterized. The aims of this study were to: (1) assemble, annotate, and conduct a comparative analysis of the complete cp. genomes of four species of *Synotis*, (2) identify similarities and differences in the structural characteristics of the cp. genomes between *Synotis* and its related species, (3) reconstruct phylogenetic relationships among *Synotis* species and its related species, (4) utilize the complete cp. genomes as a super-barcode for the identification of *Synotis* species. These results will provide valuable insights into the DNA barcoding and phylogenetics of genus *Synotis*.

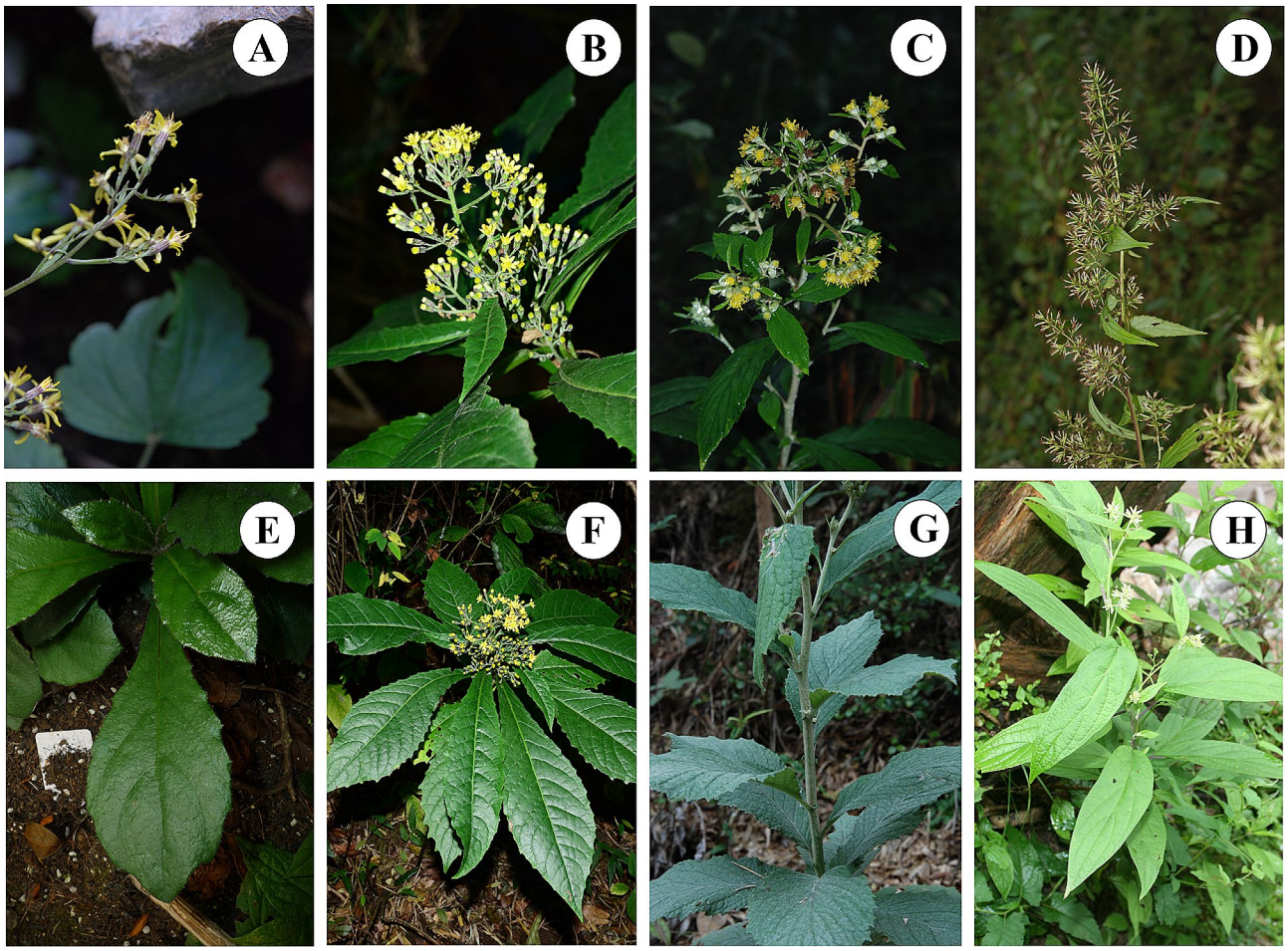

Fig. 1 Images of *Synotis cavaleriei*, *S. duclouxii*, *S. nagensium*, and *S. erythropappa*. (**A**, **B**,**C**, **D**) Inflorescence of plant growing in natural habitat; (**E**, **F**,**G**, **H**) morphology of leaf. ((**A**, **E**) *S. cavaleriei*, voucher LXF0170, SCNU, Butuo county, Liangshan prefecture, Sichuan province, China). ((**B**, **F**) *S. duclouxii*, voucher LXF0219, SCNU, Butuo county, Liangshan prefecture, Sichuan province, China). ((**C**, **G**) *S. nagensium*, voucher FZX5549, SCNU, Tongchuan District, Dazhou City, Sichuan Province, China). ((**D**, **H**) *S. erythropappa*, voucher FZX1038, SCNU, Puge county, Liangshan prefecture, Sichuan Province, China). All photographs by Ming Tang except B and F provides by Yi Yang

Results

Cp genome features of the four *Synotis* species

The gene map of the circular complete cp. genomes of four *Synotis* species is shown in Fig. 2. The GenBank numbers, GC content, gene numbers, and lengths of cp. genomes are given in Table 1. The cp. genome lengths of *S. cavaleriei*, *S. duclouxii*, *S. nagensium*, and *S. erythropappa* were 151,341 bp, 151,305 bp, 151,298 bp, and 151,228 bp, respectively (Table 1). These cp. genomes had a typical quadripartite circular structure, containing a LSC region (83,288 to 83,399 bp), an SSC region (18,262 to 18,287 bp) and two IR regions (24,837 to 24,842 bp) (Table 1; Fig. 2). The overall GC content was 37.4%, with the IR regions displaying significantly higher GC content (43.0%) compared to the LSC region (35.6%) and SSC region (30.6–30.7%).

Each of the four cp genomes contained 134 genes, including 87 protein-coding, 37 tRNA, 8 rRNA genes, and 2 pseudogenes (*ycf1* and *rps19*, Table 1). The cp genes of the four *Synotis* species can be divided into four types: named photosynthesis-related genes, self-replication-related genes, other genes and genes of unknown function. Among these genes, 7 protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7*, *ycf15*, and *ycf2*), 7 tRNAs (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*), and all 4 rRNA genes (*rrn16*, *rrn23*, *rrn4.5* and *rrn5*) were duplicated in the IR regions (Table 1). In total, there were 19 intron-containing genes (Table 1). 16 genes (*ndhA*, *ndhB*, *petB*, *petD*, *atpF*, *rbcL*, *rpl16*, *rpl2*, *rps16*, *rpoC1*, *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA* and *trnV-UAC*) contained only 1 intron and 3 genes (*clpP*, *rps12* and *ycf3*) contained 2 introns (Table 1). The *rps12* gene was trans-splicing, with the 5' end located in the LSC region and 3' end located in the IR regions. The majority of protein coding genes had the standard ATG start codon, but some genes had alternative start codons, such as ACG in *psbL*

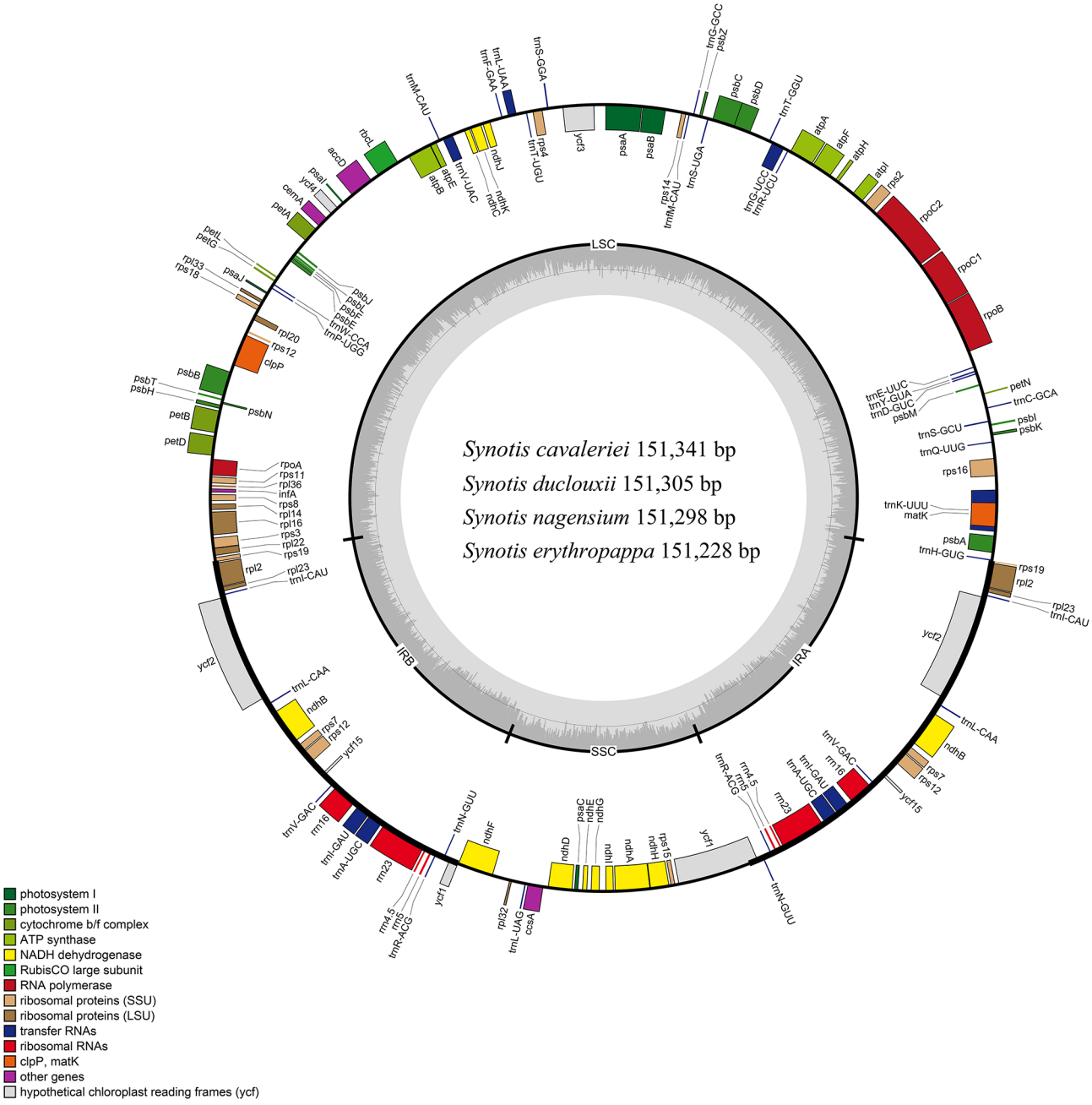

Fig. 2 Gene maps of the cp. genomes of *Synotis* species. Genes inside of the circle are transcribed clockwise and those on the outside are transcribed counter-clockwise

and *ndhD*. Four cp genome sequences of the species of *Synotis* were submitted to NCBI (accession number: OM912601, OM912602, OM912603, OQ985056).

SSRs analysis

The number and types of SSRs were generally similar among the cp. genomes of four *Synotis* species (Fig. 3). The four sequenced plastomes contained 195 SSRs, including 111 Mononucleotide repeats (57%), 16 Dinucleotide repeats (8%), 24 Trinucleotide repeats (12%), 43 Tetranucleotide repeats (22%), and 1 Pentanucleotide repeats (1%) (Fig. 3A). The most common type of SSRs were Mononucleotide repeats composed of A/T bases (Fig. 3B). Specifically, there were 52 SSRs in *S. cavaleriei*, including 31 Mononucleotides, 4 Dinucleotides, 6 Trinucleotides, and 11 Tetranucleotides, 48 SSRs in *S. duclouxii*, including 27 Mononucleotides, 4 Dinucleotides, 6 Trinucleotides, and 11 Tetranucleotides, 45 SSRs in *S. nagensium*, including 24 Mononucleotides, 4 Dinucleotides, 6 Trinucleotides, 11 Tetranucleotides, and 1

Table 1 The basic cp. Genomes information of four *Synotis* species

Characteristics	*S. cavaleriei*	*S. duclouxii*	*S. nagensium*	*S. erythropappa*
GenBank accession	OM912601	OM912602	OM912603	OQ985056
Genome size (bp)	151,341	151,305	151,298	151,228
LSC size (bp)	83,399	83,334	83,343	83,288
SSC size (bp)	18,268	18,287	18,277	18,262
IRa/IRb size (bp)	24,837	24,842	24,839	24,839
Number of genes	134	134	134	134
Protein-coding genes	87	87	87	87
Pseudogene	2	2	2	2
tRNA genes	37	37	37	37
rRNA genes	8	8	8	8
Overall GC content (%)	37.4	37.4	37.4	37.4
GC content in LSC (%)	35.6	35.6	35.6	35.6
GC content in IRa/IRb (%)	43.0	43.0	43.0	43.0
GC content in SSC (%)	30.6	30.6	30.7	30.6

Pentanucleotide, 50 SSRs in *S. erythropappa*, including 29 Mononucleotides, 4 Dinucleotides, 6 Trinucleotides, and 11 Tetranucleotides (Fig. 3C).

Structure comparison among the cp genomes of *Synotis* and its related species

The boundaries of the IR/SC were comprehensively compared (Fig. 4) in four species of Synotis and their related species. The genes located at these boundaries included *rps19*, *rpl2*, *ndhF*, *ycf1*, and *trnH* genes. The length of the IR regions ranged from 24,821 to 24,852 bp in *Synotis* and its related species, with some degree of expansion. The cp. genomes of these species were relatively conserved, except for one major divergences in *Senecio vulgaris*. In the cp. genomes of *Ligularia jaluensis*, *Sinosenecio oldhamianus*, *S. cavaleriei*, *S. duclouxii*, *S. nagensium* and *S. erythropappa*, the *ycf1* gene crossed the SSC/IRa boundary with the larger part located in the SSC region, and the *ndhF* gene located at the IRb/SSC boundary. However, in *S. vulgaris*, the *ycf1* gene crossed the IRb/SSC boundary, with 4,494 bp located in SSC region, and the *ndhF* gene was near the SSC/IRa boundary. All species contained a functional copy of the *rps19* gene at the LSC/IRb junction and a pseudo copy (*rps19Ψ*) at the IRa/LSC junction. The length of the *rps19* gene was conserved at 279 bp. The *rpl2* gene was present completely in the IRb region and was located away from the junction of LSC/IRb, with a length ranging from 113 to 119 bp. The *trnH* gene was commonly located at the LSC region, with variable gap

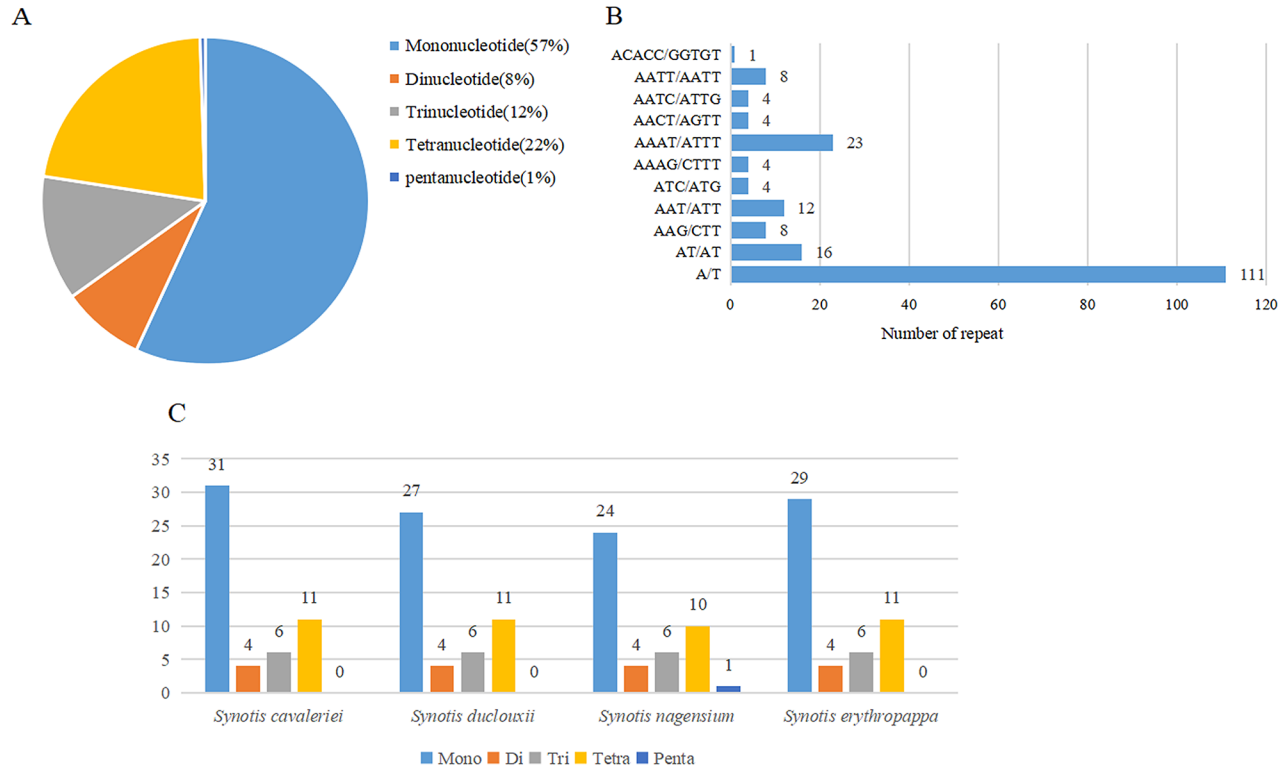

Fig. 3 Analysis of SSRs in the four newly sequenced *Synotis* cp. genomes. **(A)** Frequencies of identified SSR types in all four plastomes, **(B)** Number of different SSR motifs, (**C)** Number of SSR repeat types

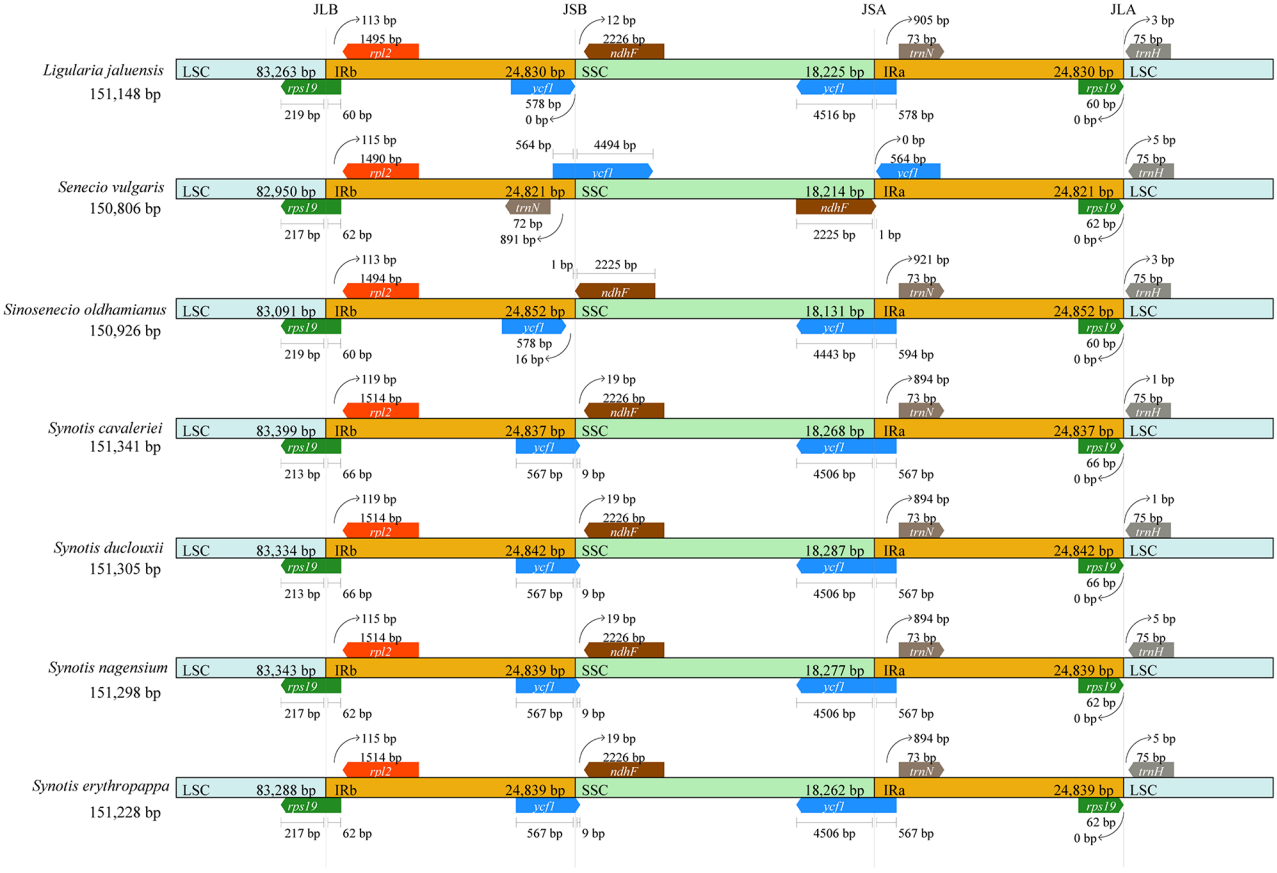

Fig. 4 Comparison of borders in the LSC, SSC and IR regions of the seven *Senecioneae* cp. genomes. Genes are denoted by colored boxes. The gaps between the genes and the boundaries are indicated by the base lengths

lengths observed in the four species of *Synotis* and its related species.

With reference to *Ligularia jaluensis*, the structural differences among cp. genomes of the four *Synotis* species and its related species were compared by mVISTA (Fig. 5). The LSC and SSC regions were more significantly divergent than the two IR regions. Additionally, the non-coding regions showed greater variability than the coding regions.

Codon usage patterns in *Synotis* cp genomes

The RSCU values were calculated for the cp. genomes of *Synotis cavaleriei*, *S. duclouxii*, *S. nagensium* and *S. erythropappa* based on the protein-coding sequences (Table 2; Fig. 6). The total sequence size coding for protein genes was 77,616–78,381 bp in the cp. genomes of four species. These protein sequences encoded 22,424–22,679 codons. The cp. genomes of the four *Synotis* species contained 64 codons. Among these, 61 codons encoded 20 amino acids, and the other 3 were termination codons. Among these amino acids, Leucine was the most abundant amino acid in the cp. genomes of four *Synotis* species, whereas cysteine was the least (Table 2). The RSCU values of all codons are shown in Fig. 6. There were 32 codons with RSCU values less than 1 (RSCU<1), which showed a lower usage frequency than expected. A total of 30 codons had RSCU values greater than 1 (RSCU>1). Both methionine (Met) and tryptophan (Trp) exhibit no codon bias and have RSCU values of 1.

Nucleotide diversity analysis

The cp. genomes contain numerous nucleotides. It can be used as the valuable DNA barcoding to resolve the phylogenetic relationships of closely related species or genera. In this study, highly variable regions were identified in *Synotis* and its related species using DnaSP (Fig. 7). Among the seven species, polymorphism information (Pi) values ranged from 0.02232 (*petN* gene) to 0.03 (*ccsA-psaC* region). The Pi analyses revealed that the IR regions displayed significantly lower variation compared to the LSC and SSC regions (Fig. 7). Six mutational hotspots had remarkably higher Pi values (>0.022), including three genes (*ycf1*, *ccsA* and *petN*) and three intergenic regions (*trnC-GCA-petN*, *ccsA-psaC* and *trnE-UUC-rpoB*).

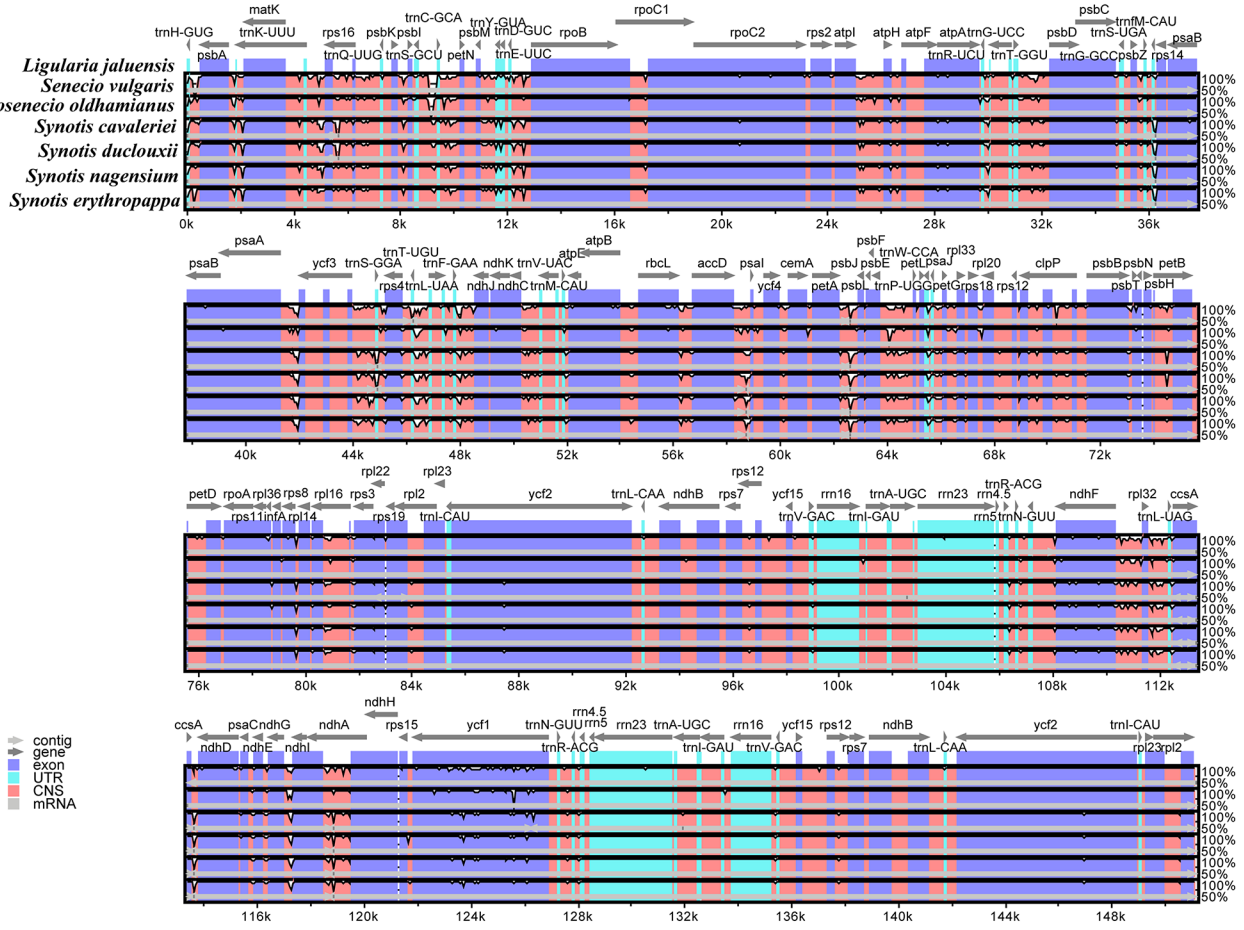

Fig. 5 Sequence alignment of seven *Senecioneae* cp. genomes using the mVISTA program with *Ligularia jaluensis* as a reference. X-axis: the coordinates in the cp. genome, Y-axis: percent identity within 50–100%. The transcriptional direction of genes indicated by grey arrows

Phylogenetic analysis

Based on the analysis of the complete cp. genome sequences, the Maximum likelihood (ML) phylogenetic tree of 42 Asteraceae species was constructed, with *Anthriscus cerefolium* (Apiaceae) and *Kalopanax septemlobus* (Araliaceae) used as outgroups (Fig. 8). In our sampling of Asteraceae, the species formed a monophyletic group and were grouped into 5 monophyletic clades, corresponding to the 5 subfamilies of Asteraceae (Asteroideae, Cichorioideae, Gymnarrhenoideae, Carduoideae and Pertyoideae). In the subfamily Asteroideae, 6 subgroups were separated, corresponding to the tribe Senecioneae, Astereae, Gnaphalieae, Anthemideae, Helenieae and Inuleae, respectively. The tribe Senecioneae contains two clades: subtribe Tussilagininae and Senecioninae. The subtribe Tussilagininae consists of 13 species in 8 genera, including *Gynoxys* Cass., *Nordenstamia* Lundin, *Roldana* La Llave, *Telanthophora* H.Rob. & Brettell, *Arnoglossum* Raf., *Sinosenecio* B.Nord., *Ligularia* Cass. and *Petasites* Mill. The subtribe Senecioninae consists of 11 species from 4 genera, including *Senecio* L., *Crassocephalum* Moench, *Dendrosenecio* (Hauman ex Hedberg) B.Nord. and *Synotis*. From the perspective of complete cp. genomes, the species of *S. cavaleriei*, *S. duclouxii*, *S. nagensium* and *S. erythropappa* were sister species, supported by a 100 bootstrap value. Among the four *Synotis* species, the species of *S. cavaleriei* clustered with *S. duclouxii*, and *S. nagensium* clustered with *S. erythropappa*. The genus of *Synotis* was phylogenetically close to *Senecio*, *Crassocephalum* and *Dendrosenecio*.

Molecular age estimation

Divergence time estimates of *Synotis* species based on the complete cp. genomes were shown in Fig. 9. The molecular age estimation suggested that *Synotis* originated at 3.81 Mya. The divergence time between the *S. cavaleriei* and *S. duclouxii* was estimated at 0.87 Mya. The divergence time between the *S. nagensium* and *S. erythropappa* was estimated at 2.78 Mya.

Table 2 Codon usage and codon-anticodon recognition patterns of four *Synotis* species

Amino acid	Symbol	Codon	Numbers and RSCU			
			S. cavaleriei	S. duclouxii	S.nagensium	S. erythropappa
*	Ter	UAA	47/1.7625	47/1.7625	47/1.7625	47/1.7625
*	Ter	UAG	18/0.675	18/0.675	18/0.675	18/0.675
*	Ter	UGA	15/0.5625	15/0.5625	15/0.5625	15/0.5625
A	Ala	GCA	352/1.1043	362/1.1173	363/1.1178	362/1.1156
A	Ala	GCC	203/0.6369	207/0.6389	210/0.6467	208/0.641
A	Ala	GCG	144/0.4518	145/0.4475	145/0.4465	147/0.453
A	Ala	GCU	576/1.8071	582/1.7963	581/1.7891	581/1.7904
C	Cys	UGC	69/0.5542	70/0.5578	70/0.5578	69/0.552
C	Cys	UGU	180/1.4458	181/1.4422	181/1.4422	181/1.448
D	Asp	GAC	180/0.4114	182/0.4113	183/0.4131	184/0.4144
D	Asp	GAU	695/1.5886	703/1.5887	703/1.5869	704/1.5856
E	Glu	GAA	879/1.509	882/1.509	885/1.5128	883/1.512
E	Glu	GAG	286/0.491	287/0.491	285/0.4872	285/0.488
F	Phe	UUC	409/0.6472	414/0.6454	414/0.6444	414/0.6444
F	Phe	UUU	855/1.3528	869/1.3546	871/1.3556	871/1.3556
G	Gly	GGA	594/1.5389	602/1.5347	602/1.5377	601/1.5361
G	Gly	GGC	175/0.4534	180/0.4589	180/0.4598	181/0.4626
G	Gly	GGG	259/0.671	260/0.6628	257/0.6564	259/0.662
G	Gly	GGU	516/1.3368	527/1.3435	527/1.3461	524/1.3393
H	His	CAC	126/0.4961	129/0.5	129/0.499	128/0.4961
H	His	CAU	382/1.5039	387/1.5	388/1.501	388/1.5039
I	Ile	AUA	626/0.9812	628/0.9782	632/0.9829	633/0.9844
I	Ile	AUC	351/0.5502	351/0.5467	352/0.5474	351/0.5459
I	Ile	AUU	937/1.4687	947/1.4751	945/1.4697	945/1.4697
K	Lys	AAA	884/1.5021	889/1.5004	887/1.4996	889/1.5004
K	Lys	AAG	293/0.4979	296/0.4996	296/0.5004	296/0.4996
L	Leu	CUA	323/0.8095	324/0.8013	323/0.7995	321/0.7939
L	Leu	CUC	141/0.3534	140/0.3462	141/0.349	142/0.3512
L	Leu	CUG	145/0.3634	146/0.3611	147/0.3639	148/0.366
L	Leu	CUU	521/1.3058	525/1.2984	525/1.2995	524/1.296
L	Leu	UUA	761/1.9073	778/1.9242	775/1.9183	777/1.9217
L	Leu	UUG	503/1.2607	513/1.2688	513/1.2698	514/1.2712
M	Met	AUG	544/1	550/1	551/1	550/1
N	Asn	AAC	233/0.4283	234/0.4251	237/0.4301	236/0.4279
N	Asn	AAU	855/1.5717	867/1.5749	865/1.5699	867/1.5721
P	Pro	CCA	275/1.1752	278/1.1767	277/1.175	277/1.175
P	Pro	CCC	170/0.7265	170/0.7196	170/0.7211	169/0.7169
P	Pro	CCG	125/0.5342	128/0.5418	129/0.5472	130/0.5514
P	Pro	CCU	366/1.5641	369/1.5619	367/1.5567	367/1.5567
Q	Gln	CAA	625/1.5432	626/1.5381	626/1.5381	625/1.5375
Q	Gln	CAG	185/0.4568	188/0.4619	188/0.4619	188/0.4625
R	Arg	AGA	414/1.8455	417/1.8465	415/1.839	413/1.8315
R	Arg	AGG	139/0.6196	139/0.6155	139/0.616	139/0.6164
R	Arg	CGA	297/1.3239	300/1.3284	301/1.3338	301/1.3348
R	Arg	CGC	93/0.4146	93/0.4118	91/0.4032	93/0.4124
R	Arg	CGG	102/0.4547	101/0.4472	101/0.4476	101/0.4479
R	Arg	CGU	301/1.3418	305/1.3506	307/1.3604	306/1.357
S	Ser	AGC	98/0.3513	99/0.3511	98/0.3475	98/0.3477
S	Ser	AGU	358/1.2832	360/1.2766	359/1.273	358/1.2703
S	Ser	UCA	349/1.2509	354/1.2553	354/1.2553	355/1.2596
S	Ser	UCC	240/0.8602	243/0.8617	243/0.8617	243/0.8622
S	Ser	UCG	128/0.4588	130/0.461	132/0.4681	131/0.4648

Table 2 (continued)

Amino acid	Symbol	Codon	Numbers and RSCU			
			S. cavaleriei	S. duclouxii	S.nagensium	S. erythropappa
S	Ser	UCU	501/1.7957	506/1.7943	506/1.7943	506/1.7954
T	Thr	ACA	348/1.2395	354/1.2443	357/1.2559	356/1.2513
T	Thr	ACC	215/0.7658	216/0.7592	213/0.7493	214/0.7522
T	Thr	ACG	110/0.3918	111/0.3902	109/0.3835	110/0.3866
T	Thr	ACU	450/1.6028	457/1.6063	458/1.6113	458/1.6098
V	Val	GUA	472/1.5375	476/1.533	473/1.5233	472/1.5189
V	Val	GUC	143/0.4658	144/0.4638	145/0.467	147/0.473
V	Val	GUG	164/0.5342	165/0.5314	166/0.5346	165/0.531
V	Val	GUU	449/1.4625	457/1.4718	458/1.475	459/1.4771
W	Trp	UGG	383/1	400/1	400/1	400/1
Y	Tyr	UAC	160/0.3778	160/0.3738	160/0.3747	161/0.3766
Y	Tyr	UAU	687/1.6222	696/1.6262	694/1.6253	694/1.6234

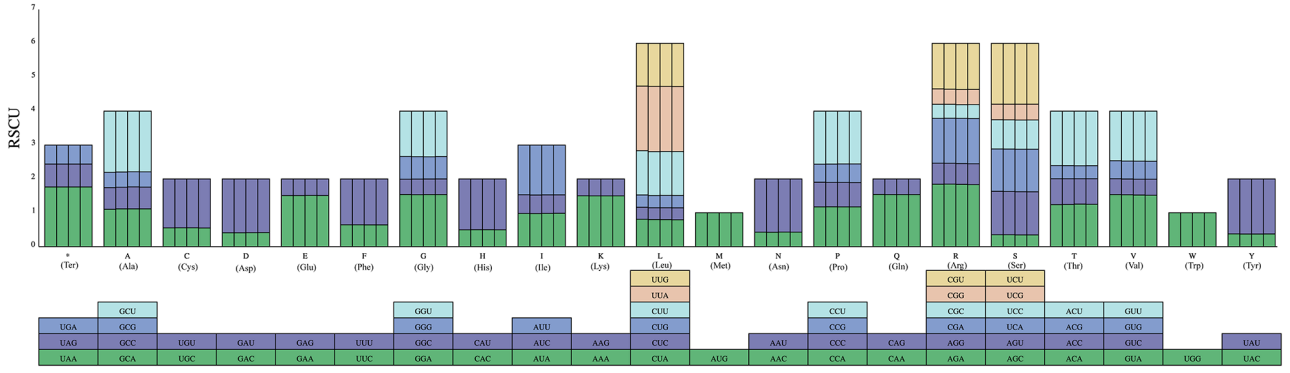

Fig. 6 Codon content of 20 amino acids and stop codons in the protein-coding genes of the cp. genomes of the four *Synotis* species. Each bar represents a species, from left to right: *S. cavaleriei*, *S. duclouxii*, *S. nagensium*, and *S. erythropappa*

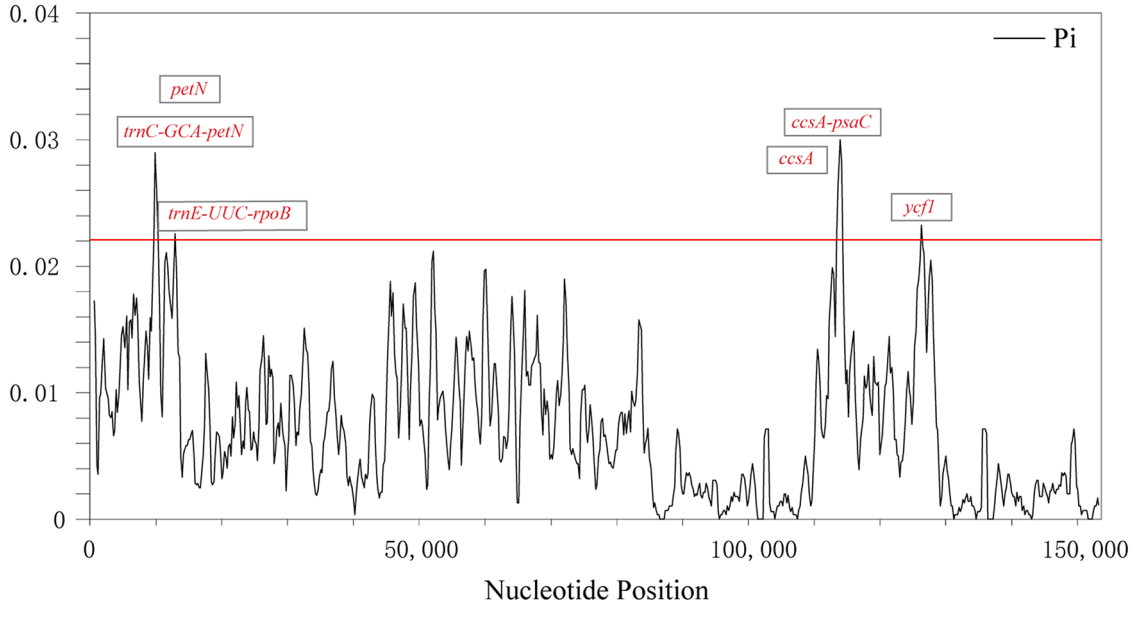

Fig. 7 Nucleotide diversity (Pi) values among the seven Senecioneae species. X-axis: position of the midpoint of a window, Y-axis: Pi value, Pi: polymorphism information

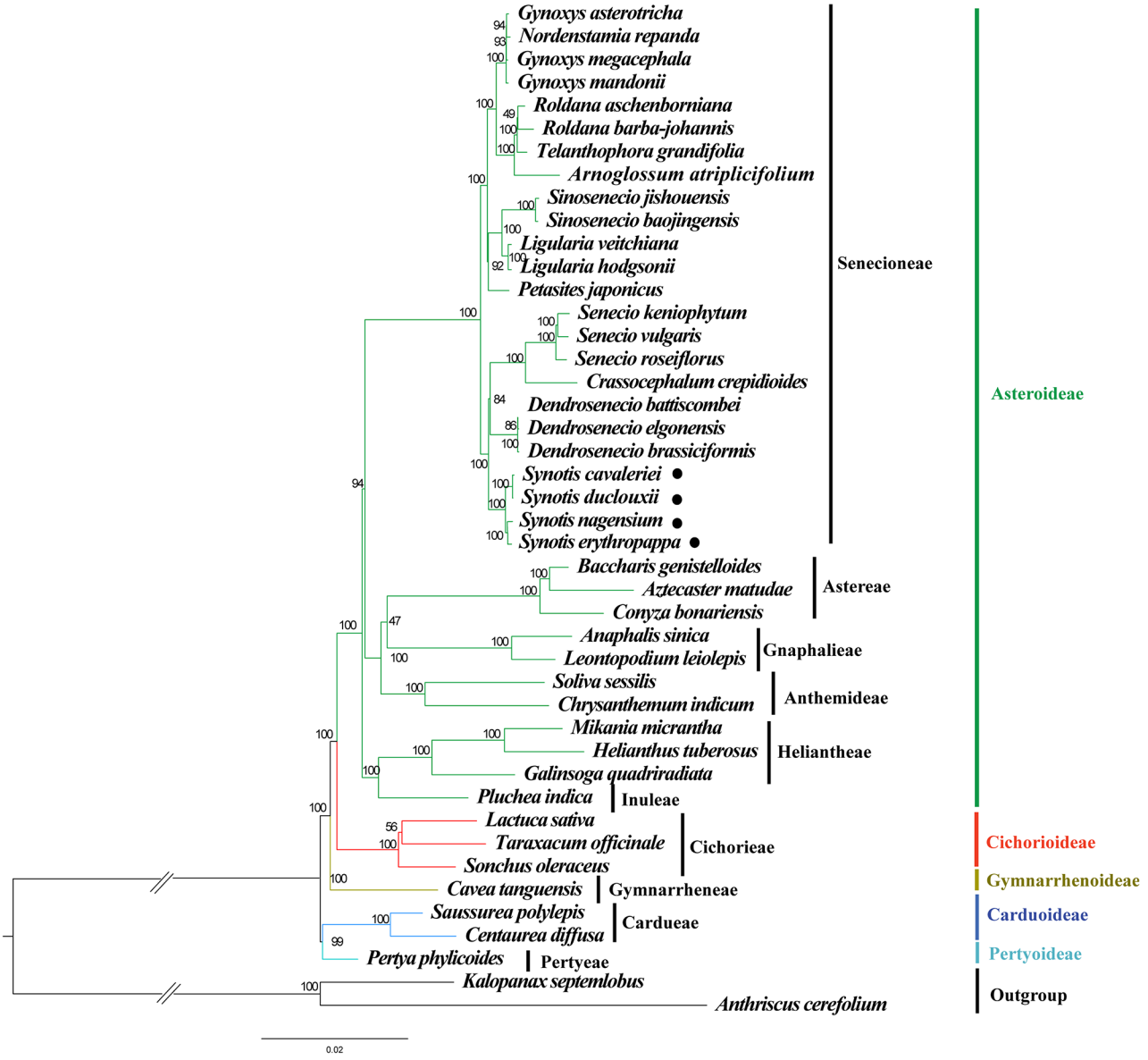

Fig. 8 Phylogenetic tree was reconstructed using ML method based on complete cp. genomes of the Asteraceae. Numbers at nodes are bootstrap support values. The phylogenetic position of *Synotis* species is marked with black circles

Discussion

Cp genomes structure analysis

In this study, the cp. genomes of four *Synotis* species were sequenced, assembled, and annotated (Fig. 2; Tables 1 and 3). This is the first report of complete cp. genomes from the genus *Synotis*. The complete cp. genomes also exhibited a highly conserved characteristic in terms of genomic numbers, orders, structures, GC contents and intron numbers. The length of these cp. genomes ranged from 151,228 (*S. erythropappa*) to 151,341 bp (*S. cavaleriei*). They are similar to other cp. genomes of tribe Senecioneae, such as *Cremanthodium rhodocephalum* [55], *Ligularia veitchiana* (151,253 bp) [25], *Parasenecio palmatisectus* (151,263 bp) [56], *Sinosenecio albonervius* (151,224 bp) [19], and *Senecio roseiflorus* (151,228 bp) [57]. Among three series within *Synotis* in this study, the length of the cp. genome of ser. *Hieraciifoliae* is the longest. The cp. genomes of four *Synotis* species encoded 134 genes, which is similar to the cp. genomes of *Ligularia* [25], *Aster* [33, 34], *Gynura* [42], and *Sinosenecio* [19]. Like most other species of Asteraceae, pseudogenes *ycf1* and *rps19* were also detected [19, 25, 30, 32, 35, 43, 56]. The ACG start codons of *psbL* and *ndhD* genes were also detected in angiosperms [38, 58–60]. The *clpP*, *rps12*, and *ycf3* genes included two introns, while the other genes included one intron. All these features are consistent with the cp. genomes of most family Asteraceae [22, 25, 43, 56]. The total GC contents of the complete

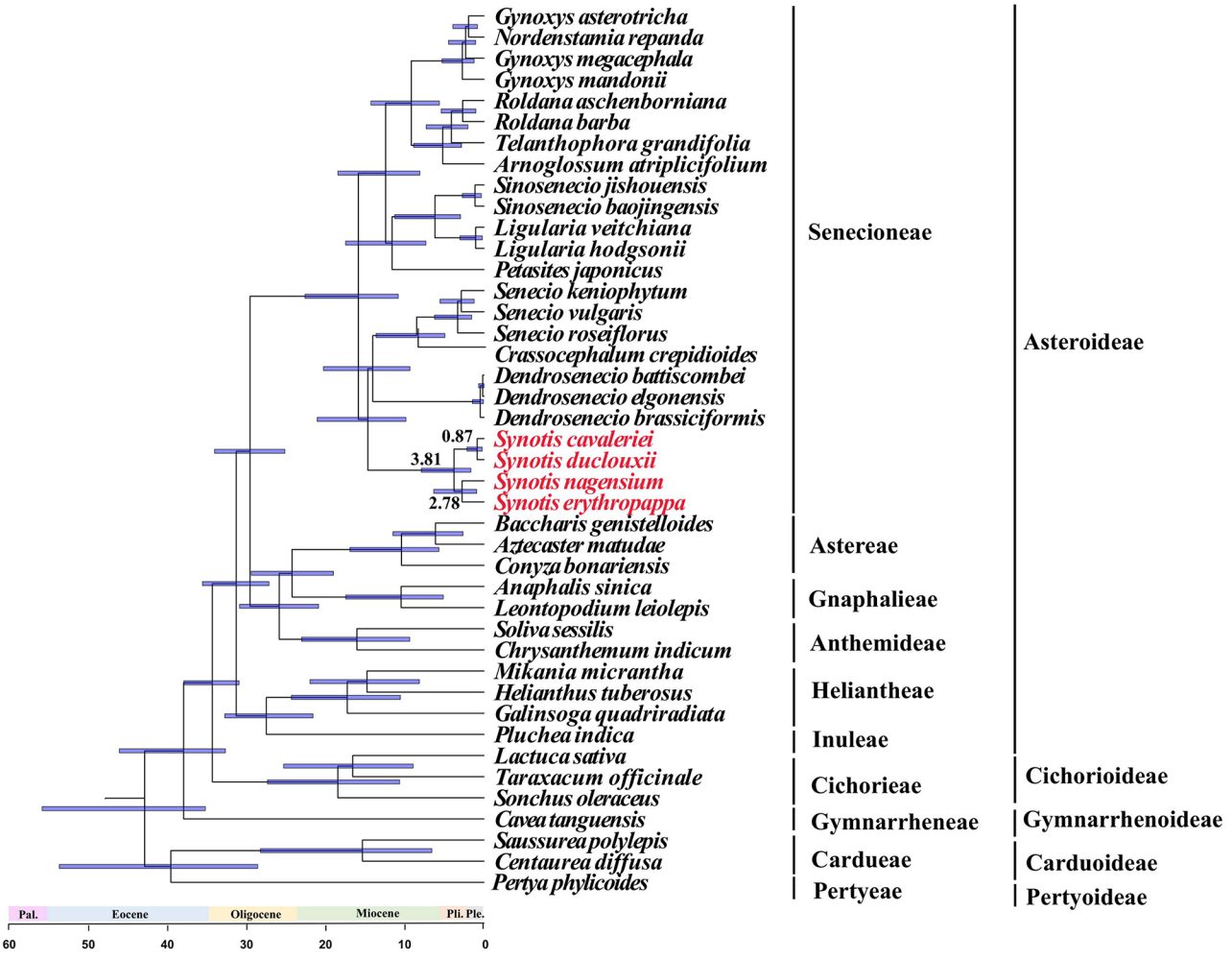

Fig. 9 Divergence time estimates of *Synotis* based on complete cp. genomes. The numbers above or below branches represent median divergence time estimates. Pal., Pli. and Ple. indicate Paleocene, Pliocene and Pleistocene, respectively. The *Synotis* species were marked in red

cp. genomes among the three series of the *Synotis* was 37.4% (Table 1), which is similar to previous study [19, 22, 25, 37, 55]. Among the LSC, SSC, and IR regions, the IR regions had the highest GC contents (43.0%), followed by the LSC (35.6%) and SSC regions (30.6 to 30.7%). The IR regions had the highest GC contents among the four regions, which is likely due to the presence of two copies of rRNAs (*rrna4.5*, *rrna5*, *rrna23*, and *rrna16*) within this region [22, 43].

Repeat sequences analysis

SSRs, also called microsatellites, are widely utilized as valuable molecular markers in species identification and phylogenetic studies due to their high substitution rates [61]. In this study, a total of 195 SSRs were identified in the complete cp. genomes of four *Synotis* species, 111 were found to be Mononucleotide repeats, constituting the majority (57%) of all SSRs (Fig. 3A). The most common SSRs are Mononucleotide repeats composed of A/T (Fig. 3B). It is worth noting that the number of SSRs varied across different species (Fig. 3C). These findings are consistent with previous reports within the Asteraceae family, such as *Artemisia* [18, 38]. In this study, we analyzed the number and composition of SSRs in the cp. genomes of four *Synotis* species. It will serve as valuable references for further research on molecular markers and population genetics of *Synotis* and the tribe Senecioneae.

Genomes comparison and nucleotide diversity

Comparative analysis of the complete cp. genomes provides a pivotal reference in understanding genome evolution and evolutionary relationships within the family Asteraceae [18, 19, 22, 23, 56]. Through the comparative analysis of the IR boundary regions among *Synotis* and its related species, some expansion or contractions were detected, with the IR regions ranging from 24,821 to 24,852 bp (Fig. 4). Similar large-scale IR expansion had also been reported in other Asteraceae, such as *Saussurea* [17], *Aster* [35], *Xanthium* [46]. Furthermore, the SSC and LSC regions showed higher sequence divergence

Table 3 The basic cp. Genomes information of four *Synotis* species

Category	Gene group	Gene name
Photosynthesis	Subunits of photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Subunits of photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Subunits of NADH dehydrogenase	*ndhA**, *ndhB**(2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Subunits of cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*
	Subunits of ATP synthase	*atpA*, *atpB*, *atpE*, *atpF**, *atpH*, *atpI*
	Large subunit of rubisco	*rbcL**
Self-replication	Proteins of large ribosomal subunit	*rpl14*, *rpl16**, *rpl2**(2), *rpl20*, *rpl22*, *rpl23*(2), *rpl32*, *rpl33*, *rpl36*
	Proteins of small ribosomal subunit	#*rps19*, *rps11*, *rps12***(2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*, *rps2*, *rps3*, *rps4*, *rps7*(2), *rps8*
	Subunits of RNA polymerase	*rpoA*, *rpoB*, *rpoC1**, *rpoC2*
	Ribosomal RNAs	*rrn16*(2), *rrn23*(2), *rrn4.5*(2), *rrn5*(2)
	Transfer RNAs	*trnA-UGC**(2), *trnC-GCA*, *trnD-GUC*, *trnE-UUC*, *trnF-GAA*, *trnG-GCC*, *trnG-UCC**, *trnH-GUG*, *trnI-CAU*(2), *trnI-GAU**(2), *trnK-UUU**, trn*L-CAA*(2), *trnL-UAA**, *trnL-UAG*, *trnM-CAU*, *trnN-GUU*(2), *trnP-UGG*, *trnQ-UUG*, *trnR-ACG*(2), *trnR-UCU*, *trnS-GCU*, *trnS-GGA*, *trnS-UGA*, trn*T-GGU*, *trnT-UGU*, *trnV-GAC*(2), *trnV-UAC**, *trnW-CCA*, *trnY-GUA*, *trnfM-CAU*
Other genes	Maturase	*matK*
	Protease	*clpP***
	Envelope membrane protein	*cemA*
	Acetyl-CoA carboxylase	*accD*
	c-type cytochrome synthesis gene	*ccsA*
	Translation initiation factor	*infA*
Genes of unknown function	Conserved hypothetical chloroplast ORF	#*ycf1*, *ycf1*, *ycf15*(2), *ycf2*(2), *ycf3***, *ycf4*

Notes: *: Gene with one introns, **: Gene with two introns, #: Pseudogene, (2): Gene with two copies

than the IR regions, and the non-coding region was more divergence than the coding region (Fig. 5), the same result was also obtained in Senecioneae [19, 44, 57]. It should be noted that mutation hotspots can serve as specific DNA barcodes for species identification [62]. In this study, six highly variable regions (Pi > 0.022) were identified, including *trnC-petN*, *ccsA-psaC*, *trnE-rpoB*, *ycf1*, *ccsA* and *petN*. These variable regions could be utilized as potential molecular barcodes for species identification and phylogenetic research in genus *Synotis* and tribe Senecioneae.

Phylogenetic relationship, super barcode and molecular dating analysis

In recent years, several molecular phylogenetic analyses have been conducted to resolve the phylogenetic relationships of genus *Synotis*. Liu et al. (2006) studied the radiation and diversification within the *Ligularia-Cremanthodium-Parasenecio* complex in the Qinghai-Tibetan Plateau, they used *Synotis lucorum* as the outgroup and inferred a phylogenetic tree of the LCP complex using ITS, *ndhC-F* and *trnL-F* [63]. Pelser et al. (2007) constructed the phylogenetic relationships for the Senecioneae using ITS data. The results showed that the species of *Synotis* was nested within the *Cissampelopssis-Crassocephalum* clade [51]. However, these studies were based on small taxon sampling and small fragments of barcoding markers. Tang (2014) reconstructed the Phylogenetic tree of the *Synotis* using ITS and seven cpDNA sequence regions. In his analyses, the monophyly of *Synotis* is strongly supported and the species of *Senecio kumaonensis* is also included [6]. Although some studies of phylogenetic relationships have been reported using limited nuclear and cpDNA fragments in the above studies, the research progress of genus *Synotis* at the complete cp. genome-scale level is relatively slow. The result indicated that all species were grouped into five monophyletic clades (Fig. 8), corresponding to the five subfamilies of Asteraceae, which was consistent with the previous study [64]. In this study, the phylogenetic trees among the three series of *Synotis* were constructed using complete cp. genomes for the first time. In the ML tree, the *Synotis* was a branch of family Asteraceae, the four *Synotis* species are sister groups within tribe Senecioneae and subfamily Asteroideae, and the bootstrap support values for these were 100. Our study also confirms the monophyly of *Synotis.* In the tribe Senecioneae, two subgroups were separated, corresponding to the subtribe Tussilagininae and Senecioninae, respectively. The genus *Synotis* belongs to the Senecioninae subtribe, which is consistent with the backbone of Senecioneae in previous studies [47, 51, 64].

The complete cp. genomes have been widely served as super barcode for species identification and classification [25–29]. In this study, the Senecioneae contains two clades. Within subtribe Senecioninae, a sub-clade consisting of four *Synotis* species was identified with high bootstrap support values (bootstrap = 100) (Fig. 8). Four

Synotis species were discriminated completely based on complete cp. genome sequences. This topology shows the separation of *Synotis* at intra-species level, which suggested that the complete cp. genomes of *Synotis* might be a super barcode for species identification.

Previous studies have estimated the divergence time of the Senecioneae tribe during the Eocene period using various molecular markers. For instance, Panero and Crozier (2016) used a few cp. genomes markers [65], Han et al. (2019) utilized complete cp. genomes [42], and Mandel et al. (2019) employed nuclear phylogenomics data [66]. Huang et al. (2016) estimated the divergence time of the tribe Senecioneae during the early Miocene based on multiple nuclear genes [67]. In this study, the divergence time of the tribes Senecioneae was early Oligocene period. This study provides further support and traces the divergence time of the Senecioneae tribe.

In this study, it is an efficient attempt to utilize the complete genomes as super-barcode for the identification of genus *Synotis*. Our research provides a foundation for developing a reliable and efficient method for species identification of *Synotis.* In the future, it is necessary to further investigate and verify this assumption. In addition, *Synotis* is a large genus with ca. 60 species [1], and this study only analyzes the cp. genome sequences of four species. Therefore, a more comprehensive analysis of cp. genomes and a broader taxon sampling are essential in order to fully develop the super-barcode approach. The development of such a method will greatly contribute to biodiversity assessments, conservation efforts, and various fields of plant research.

Conclusions

This study reported the complete cp. genomes from four *Synotis* species, ranging from 151,228 (*S. erythropappa*) to 151,341 bp (*S. cavaleriei*). The structure and composition, SSRs, codon usage of the complete cp. genomes were highly conserved. The complete cp. genomes of four *Synotis* species encoded 134 genes, including 87 protein-coding genes, 37 tRNA genes, 8 rRNA genes, and 2 Pseudogenes (*ycf1* and *rps19*). Phylogenetic analysis showed that four species were clustered into a monophyletic group, and they were close to the genus of *Senecio*, *Crassocephalum* and *Dendrosenecio* in tribe Senecioneae. The ML tree showed that the complete cp. genomes could be used as a super-barcode for the identification of *Synotis* species.

Materials and methods

Plant materials and DNA extraction

Fresh leaves of *Synotis cavaleriei* and *S. duclouxii* were collected from Butuo County, Liangshan Prefecture, Sichuan Province, China. Fresh leaves of *S. nagensium* were collected from Tongchuan District, Dazhou City, Sichuan Province, China. Fresh leaves of *S. erythropappa* were collected from Puge County, Liangshan Prefecture, Sichuan Province, China. (Fig. 1). These specimens were identified by Dr. Zhixi Fu at Sichuan Normal University. The voucher specimen of *S. cavaleriei* (LXF0170), *S. duclouxii* (LXF0219), *S. nagensium* (FZX5549) and *S. erythropappa* (FZX1038) were deposited in the herbarium of Sichuan Normal University (SCNU) (Contact: Prof. Dr. Zhixi Fu, fuzx2017@sicnu.edu.cn). We achieved all required permits for the protected areas from the local governments and National Park Services. This research was carried out in compliance with the relevant laws of China. Total genomic DNA was extracted from fresh leaves using the cetyltrimethylammonium bromide (CTAB) method, following the manufacturer's instructions [68]. The Illumina Paired-End (PE) DNA Library Kit (Illumina, San Diego, CA, USA) was used to construct the DNA libraries. The genomic library was sequenced using the Illumina NovaSeq 6000 platform with a 150 bp read length (NovoGene Inc., Beijing, China).

Genomes assembly and annotation

The complete cp. genome sequences of *Synotis* were assembled using SPAdes software (v3.10.1) with default parameters [69]. The assembly quality was assessed using Bandage software to identify circular maps [70]. The cp. sequences were then annotated using PGA based on reference cp. genome sequences of *Ligularia jaluensis* (MF539931.1) [71]. The annotation results were manually inspected and adjusted using Geneious [72]. Circular maps of the cp. genomes were drawn using Organellar Genome Draw (OGDraw) (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html (Accessed August 6, 2022)) [73]. Various plastome characteristics, such as the number of genes, gene length, GC contents and intron numbers were analyzed using Geneious [72]. The obtained cp. genome sequences were deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/) with the GenBank accession numbers OM912601 (*S. cavaleriei*), OM912602 (*S. duclouxii*), OM912603 (*S. nagensium*), and OQ985056 (*S. erythropappa*).

Simple sequence repeats (SSRs) analysis

SSRs were analyzed using the MISA software (http://pgrc.ipkgatersleben.de/misa/ (Accessed January 3, 2023)) [74]. The minimum number of repeats for Mono-, Di-, Tri-, Tetra-, Penta- and Hexanucleotide motifs were set to 8, 5, 3, 3, 3, and 3, respectively.

Comparative analysis of cp genomes

Three cp. genomes from tribe Senecioneae were downloaded from GenBank to conduct comparative cp. genomes analysis. The species included in the analysis were *Ligularia jaluensis* Kom., *Senecio vulgaris* L. and

Sinosenecio oldhamianus (Maxim.) B.Nord. The boundary differences among these genomes were visualized using mVISTA in Shuffle-LAGAN mode, with the *Ligularia jaluensis* as the reference (http://genome.lbl.gov/vista/mvista/submit.shtml (Accessed August 15, 2022)) [75]. The contraction and expansion of the IR boundaries among the four regions (LSC/IRa/SSC/IRb) of the cp. genome sequences were observed using the online software IRSCOPE (https://irscope.shinyapps.io/irapp/ (Accessed January 3, 2023)) [76].

Codon usage analysis

The distribution of codon usage was studied using CodonW (http://downloads.fyxm.net/ CodonW-76666.html (Accessed February 4, 2023)) with the relative synonymous codon usage (RSCU) ratio [77]. A value of RSCU>1 indicates that the codon is used more frequently, a value of 1 indicates no codon usage preference, and a value of <1 indicates less frequent codon usage.

Genetic divergence analysis

To identify and analyze highly variable regions of the cp. genomes sequence, the sequences of seven Senecioneae cp. genomes were aligned using the MAFFT program [78]. The nucleotide diversity (Pi) was then analyzed through the sliding window using DnaSP v5 (http://www.ub.edu/dnasp/ (Accessed January 6, 2023)) [79]. The step size was set to 200 bp with an 800 bp window length.

Phylogenetic analysis

Phylogenetic analysis was inferred based on 44 complete cp. genomes, including the 4 newly sequences of *Synotis* in this study, 38 cp. genomes of family Asteraceae downloaded from the NCBI database (Table S1), and *Anthriscus cerefolium* (Apiaceae) and *Kalopanax septemlobus* (Araliaceae) as outgroup. To align the sequences, the MAFFT program was used with the auto strategy [78]. The initial alignment was further manually adjusted in BioEdit to ensure accuracy [80]. The maximum-likelihood (ML) phylogenetic tree was constructed in RaxML 8.2.10 tool at CIPRES Science Gateway web (https://www.phylo.org/ (Accessed January 7, 2023)) with 1000 bootstrap replicates [81, 82].

Estimation of divergence times

The divergence times of genus *Synotis* were estimated using BEAST2 by lognormal relaxed clock [83]. The GTP model was chosen as a substitution model. The GTR model was chosen to generate the tree. As a calibration point, the pairwise divergence time of Asteroideae was set at 37.7 Mya [66]. A total of 10,000,000 generations were run. Tracer v.1.7 was used to check the correctness [84]. Finally, the maximum clade credibility tree was calculated in TreeAnnotator v1.8.4.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10663-x.

Supplementary Material 1

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for the constructive criticism of the original manuscript. We also thank Dr. Yi Yang from Jiangxi Agricultural University provides the color images of *Synotis duclouxii*.

Author contributions

M.T. and Z.F. conceived and designed the research. X.L. and J.L. performed bioinformatic analyses. X.L., H.C. and J.L. carried out wet-lab experiments. T.L. and T.Q. contributed to data interpretation. X.L. and Z.F. wrote the manuscript. M.T. and Z.F. revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the National Natural Science Foundation of China (No. 32000158, No. 31500166, No. 31960043), the National Science & Technology Fundamental Resources Investigation Program of China (No. 2021XJKK0702), and the Foundation of Sustainable Development Research Center of Resources and Environment of Western Sichuan, Sichuan Normal University (No. 2020CXZYHJZX03). Jiangxi Provincial Key Laboratory of Conservation Biology (No. 2023SSY02081).

Data availability

All annotated chloroplast genomes have been deposited in GenBank as OM912601 (*Synotis cavaleriei*), OM912602 (*S. duclouxii*), OM912603 (*S. nagensium*), and OQ985056 (*S. erythropappa*).

Declarations

Ethics approval and consent to participate

The study was conducted the plant material that complies with relevant institutional, national, and international guidelines and legislation.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 9 July 2023 / Accepted: 25 July 2024
Published online: 07 August 2024

References

1. Chen YL, Nordenstam B, Jeffrey C. *Synotis* (C. B. Clarke) C. Jeffrey & Y. L. Chen. – In: Wu, Z. Y. and Raven, P. H, editors, Flora of China. Vol. 20–21. Science Press; Miss. Bot. Gard. Press. 2011;pp.489–505.
2. Joshi S, Shrestha K, Bajracharya DM. *Synotis managensis* (Senecioneae: Asteraceae) – a new species from Manang, central Nepal. Pleione. 2013;7:539–43.
3. Tang M, Hong Y, Yang QE. *Synotis baoshanensis* (Asteraceae), a new species from Yunnan. China Bot Stud. 2013a;54:e17.
4. Tang M, Wang LY, Yang QE. *Synotis xinningensis* (Asteraceae), a new species from Hunan. China Bot Stud. 2013b;54:e16.
5. Tang M, Wang LY, Yang QE. The identity of *Synotis cordifolia* (Asteraceae–Senecioneae). J Trop Subtrop Bot. 2013c;21:101–8.
6. Tang M. A Systematic Study of the Genus *Synotis* (Compositae-Senecioneae). PhD thesis. University of Chinese Academy of Sciences, Beijing. 2014.
7. Li Z, Zheng HL, Tang M. *Synotis panzhouensis* (Asteraceae, Senecioneae), a distinct new species with red-purple pappus from southwestern Guizhou, China. Phytokeys. 2020;166:79–86.
8. Liu YL, Zhang R, Ren C, Liu B, Zhang Y, Tang M. On the specific identity of Chinese endemic species *Synotis longipes*. Phytotaxa. 2020;472:269–76.

9. Liu YL, Zhu XX, Peng YL, Tang M. *Synotis jinshajiangensis* (Asteraceae: Senecioneae), a new species from northwestern Yunnan. China Phytotaxa. 2021;478:162–70.
10. Sennikov AN, Nuraliev MS, Kuznetsov AN, Kuznetsova SP. New national records of Asteraceae from Hoang Lien National Park, northern Vietnam. Wulfenia. 2020;27:1–9.
11. Tang M, Chen YS. *Blumea hunanensis* is a synonym of *Synotis nagensium* (Asteraceae: Senecioneae). Phytotaxa. 2021;487:149–56.
12. Zhang R, Liu YL, Tang M. Four new synonyms in *Synotis* (Asteraceae, Senecioneae). Phytotaxa. 2021;483:255–66.
13. Fan Y, Zhang R, Guo CL, Tang M. On the identity of *Synotis ionodasys* (Asteraceae: Senecioneae) from Yunnan, China. Phytotaxa. 2022;545:207–16.
14. Jeffrey C, Chen YL. Taxonomic studies on the tribe Senecioneae (Compositae) of eastern Asia. Kew Bull. 1984;39:205–446.
15. Wicke S, Schneeweiss GM, de Pamphilis CW, Muller 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.
16. Shaw J, Lickey EB, Schilling EE, Small RL. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot. 2007;94:275–88.
17. Zhang X, Deng T, Moore MJ, Ji YH, Lin N, Zhan HJ, Meng AP, Wang HC, Sun YX, Sun H. Plastome phylogenomics of *Saussurea* (Asteraceae: Cardueae). BMC Plant Biol. 2019;19:290.
18. Lan Z, Shi Y, Yin Q, Gao R, Liu C, Wang W, Tian X, Liu J, NongY, Xiang L, et al. Comparative and phylogenetic analysis of complete chloroplast genomes from five *Artemisia* species. Front. Plant Sci. 2022;13:1049209.
19. Peng JY, Zhang XS, Zhang DG, Wang Y, Deng T, Huang XH, Kuang TH, Zhou Q. Newly reported chloroplast genome of *Sinosenecio albonervius* Y. Liu & Q. E. Yang and comparative analyses with other *Sinosenecio* species. BMC Genom. 2022;23:639.
20. Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 2016;17:134.
21. Jansen RK, Ruhlman TA. Genomics Chloroplasts Mitochondria. 2012;35:103–26.
22. Shahzadi I, Mehmood F, Ali Z, Ahmed I, Mirza B. Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae. Genomics. 2020;112:1454–63.
23. Jin GZ, Li WJ, Song F, Yang L, Wen ZB, Feng Y. Comparative analysis of complete *Artemisia* subgenus *Seriphidium* (Asteraceae: Anthemideae) chloroplast genomes: insights into structural divergence and phylogenetic relationships. BMC Plant Biol. 2023;23:136.
24. Yu SH, Yang XC, Tian XY, Liu XF, Huang CP, Fu ZX. The complete chloroplast genome sequence of the monotypic and enigmatic genus *Cavea* (tribe Gymnarrheneae) and a comparison with other species in Asteraceae. J Genet. 2022;101:20.
25. Chen X, Zhou J, Cui Y, Wang Y, Duan B, Yao H. Identification of *Ligularia* herbs using the complete chloroplast genome as a super-barcode. Front Pharmacol. 2018;9:695.
26. Wang AS, Wu HW, Zhu XC, Lin JM. Species identification of *Conyza bonariensis* assisted by chloroplast genome sequencing. Front Genet. 2018;9:374.
27. Zhang ZL, Zhang Y, Song MF, Guan YH, Ma XJ. Species identification of *Dracaena* using the complete chloroplast genome as a super-barcode. Front Pharmacol. 2019;10:1441.
28. Wu L, Wu ML, Cui N, Xiang L, Li Y, Li XW, Chen SL. Plant super-barcode: a case study on genome-based identification for closely related species of *Fritillaria*. Chin Med. 2021;16:52.
29. Wang WT, Wang XW, Shi YH, Yin QG, Gao RR, Wang MY, Xiang L, Wu L. Identification of *Laportea bulbifera* using the complete chloroplast genome as a potentially effective super-barcode. J Appl Genet. 2023;64:231–45.
30. Liu Y, Huo NX, Dong LL, Wang Y, Zhang SX, Young HA, Feng XX, Gu YQ. Complete chloroplast genome sequences of mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. PLoS ONE. 2013;8:e57533.
31. Choi KS, Park SJ. The complete chloroplast genome sequence of *Aster Spathulifolius* (Asteraceae); genomic features and relationship with Asteraceae. Gene. 2015;572:214–21.
32. Park J, Shim J, Won H, Lee J. Plastid genome of *Aster altaicus* var. *Uchiyamae* Kitam., an endanger species of Korean asterids. J Species res. 2017;6:76–90.
33. Ou CZ, Feng YL, Hu YK, Tian XY, Fu ZX. Characterization of the complete chloroplast genome sequence of *Aster hersileoides* (Asteraceae, Astereae) and its phylogenetic implications. Mitochondrial DNA Part B. 2019;4:985–6.
34. Wang Q, Wang XH, Yang XC, Fu LF, Gong QJ, Fu ZX. The complete chloroplast genome of *Aster hypoleucus* (Asteraceae: Astereae): an endemic species from China. Mitochondrial DNA Part B. 2019;4:2647–8.
35. Tyagi S, Jung JA, Kim JS, Won SY. Comparative analysis of the complete chloroplast genome of mainland *Aster Spathulifolius* and other *Aster* species. Plants. 2020;9:568.
36. Zhang X, Jiang PP, Fan SJ. Characterization of the complete plastome of *Aster pekinensis* (Asteraceae), a perennial herb. Mitochondrial DNA Part B. 2021;6:1064–5.
37. Zheng T, Liu M, Wang T, Chen H. The complete chloroplast genome sequence of medicinal plant *Conyza Blinii* H. Lév. Mitochondrial DNA Part B. 2017;2:50–1.
38. Ma YP, Zhao L, Zhang WJ, Zhang YH, Xing X, Duan XX, Hu J, Harris A, Liu PL, Dai SL, et al. Origins of cultivars of *Chrysanthemum*—evidence from the chloroplast genome and nuclear *LFY* gene. J Syst Evol. 2020;58:925–44.
39. Do HDK, Jung J, Hyun J, Yoon SJ, Lim C, Park K, Kim JH. The newly developed single nucleotide polymorphism (SNP) markers for a potentially medicinal plant, *Crepidiastrum denticulatum* (Asteraceae), inferred from complete chloroplast genome data. Mol Biol Rep. 2019;46:3287–97.
40. Shen J, Zhang X, Landis JB, Zhang H, Deng T, Sun H, Wang H. Plastome evolution in *Dolomiaea* (Asteraceae, Cardueae) using phylogenomic and comparative analyses. Front Plant Sci. 2020;11:376.
41. Zhou JY, Li JJ, Peng SB, An XM. Characterization of the complete chloroplast genome of the invasive plant *Erigeron annuus* (L.) Pers. (Asterales: Asteraceae). Mitochondrial DNA Part B. 2022;7:188–90.
42. Han TY, Li MM, Li JW, Lv H, Ren BR, Chen J, Li WL. Comparison of chloroplast genomes of *Gynura* species: sequence variation, genome rearrangement and divergence studies. BMC Genom. 2019;20:791.
43. Duan N, Deng LL, Zhang Y, Shi YC, Liu BB. Comparative and phylogenetic analysis based on chloroplast genome of *Heteroplexis* (Compositae), a protected rare genus. BMC Plant Biol. 2022;22:1–10.
44. Liu XF, Han MY, Chen J, Zhang X, Chen ZY, Tang Y, Qu TM, Huang CP, Yu SH, Fu ZX. Characterization of the complete chloroplast genome sequence of Chinese endemic species of *Nouelia Insignis* (Hyalideae, Asteraceae) and its phylogenetic implications. Mitochondrial DNA Part B. 2022;7:600–2.
45. Yun S, Kim SC. Comparative plastomes and phylogenetic analysis of seven Korean endemic *Saussurea* (Asteraceae). BMC Plant Biol. 2022;22:550.
46. Raman G, Park KT, Kim JH, Park S. Characteristics of the completed chloroplast genome sequence of *Xanthium spinosum*: comparative analyses, identification of mutational hotspots and phylogenetic implications. BMC Genom. 2020;21:855.
47. Pelser PB, Kennedy AH, Tepe EJ, Shidler JB, Nordenstam B, Kadereit JW, Watson LE. Patterns and causes of incongruence between plastid and nuclear senecioneae (Asteraceae) phylogenies. Am J Bot. 2010;97:856–73.
48. Tang M, Ren C, Yang QE. Parasenecio Chenopodiifolius (Compositae–Senecioneae) is a Synotis and conspecific with S. otophylla based on evidence from morphology, cytology and ITS/ETS sequence data. Nord J Bot. 2014;32:824–35.
49. Tong TJ, Tang M, Ren C, Yang QE. *Senecio kumaonensis* (Asteraceae, Senecioneae) is a *Synotis* based on evidence from karyology and nuclear ITS/ETS sequence data. Phytotaxa. 2017;292:35–46.
50. Li HM, Lazkov GA, Illarionova ID, Tong TJ, Shao YY, Ren C. Transfer of *Senecio karelinioides* (Asteraceae Senecioneae) to *Synotis* based on evidence from morphology, karyology and ITS/ETS sequence data. Nord J Bot. 2018;32:1–12.
51. Pelser PB, Nordenstam B, Kadereit JW, Watson LE. An ITS phylogeny of tribe Senecioneae (Asteraceae) and a new delimitation of *Senecio* L. Taxon. 2007;56:1077–104.
52. Moore MJ, Dhingra A, Soltis PS, Shaw R, Farmerie WG, Folta KM, Soltis DE. Rapid and accurate pyrosequencing of angiosperm plastid genomes. BMC Plant Biol. 2006;6:17.
53. Parks M, Cronn R, Liston A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. BMC Biol. 2009;7:84.
54. Ruhsam M, Rai HS, Mathews S, Ross TG, Graham SW, Raubeson LA, Mei W, Thomas PI, Gardner MF, Ennos RA, et al. Does complete plastid genome sequencing improve species discrimination and phylogenetic resolution in Araucaria? Mol Ecol Resour. 2015;15:1067–78.
55. Zhong WH, Du XL, Wang XY, Cao L, Mu ZJ, Zhong GY. Comparative analyses of five complete chloroplast genomes from the endemic genus *Cremanthodium* (Asteraceae) in Himalayan and adjacent areas. Physiol Mol Biol Plants. 2023;29:409–20.

56. Liu XF, Luo JJ, Zhang MK, Wang Q, Liu J, Wu D, Fu ZX. Phylogenomic analysis of two species of *Parasenecio* and comparative analysis within tribe Senecioneae (Asteraceae). Diversity. 2023;15:563.
57. Gichira AW, Avoga S, Li ZZ, Hu GW, Wang QF, Chen JM. Comparative genomics of 11 complete chloroplast genomes of Senecioneae (Asteraceae) species: DNA barcodes and phylogenetics. Bot Stud. 2019;60:3–17.
58. Raman G, Park S. The complete chloroplast genome sequence of Ampelopsis: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. Front Plant Sci. 2016;7:341.
59. Gichira AW, Li Z, Saina JK, Long Z, Hu G, Gituru RW, et al. The complete chloroplast genome sequence of an endemic monotypic genus *Hagenia* (Rosaceae): structural comparative analysis, gene content and microsatellite detection. PeerJ. 2017;5:e2846.
60. Alzahrani DA, Yaradua SS, Albokhari EJ, Abba A. Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. BMC Genom. 2020;21:393.
61. Upadhyay A, Kadam US, Chacko P, Karibasappa GS. Microsatellite and RAPD analysis of grape (*Vitis* spp.) accessions and identification of duplicates/misnomers in germplasm collection. Indian J Hortic. 2010;67:8–15.
62. Li HL, Xiao WJ, Tong T, Li YL, Zhang M, Lin XX, et al. The specific DNA barcodes based on chloroplast genes for species identification of Orchidaceae plants. Sci Rep. 2021;11:1424.
63. Liu J, Wang Y, Wang A, Hideaki O, Abbott R. Radiation and diversification within the *Ligularia-Cremanthodium-Parasenecio* complex (Asteraceae) triggered by uplift of the Qinghai-Tibetan plateau. Mol Phylogenet Evol. 2006;38:31–49.
64. Fu ZX, Jiao BH, Nie B, Zhang GJ, Gao TG. A comprehensive generic-level phylogeny of the sunflower family: implications for the systematics of Chinese Asteraceae. J Syst Evol. 2016;54:416–37.
65. Panero JL, Crozier BS. Macroevolutionary dynamics in the early diversification of Asteraceae. Mol Phylogenet Eevol. 2016;99:116–32.
66. Mandel JR, Dikow RB, Siniscalchi CM, Thapa R, Watson LE, Funk VA. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. Proc Natl Acad Sci. 2019;116:14083–8.
67. Huang CH, Zhang CF, Liu M, Hu Y, Gao TG, Qi J, Ma H. Multiple polyploidization events across Asteraceae with two nested events in the early history revealed by nuclear phylogenomics. Mol Biol Evol. 2016;33:2820–35.
68. Allen G, Flores-Vergara M, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. 2006;1:2320–5.
69. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
70. Wick RR, Schultz MB, Zobel J, Holt KE, Bandage. Interactive visualisation of de novo genome assemblies. Bioinform. 2015;31:3350–2.
71. Qu XJ, Moore MJ, Li DZ, Yi TS. PGA: A software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 2019;15:50.
72. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinform. 2012;28:1647–9.
73. Greiner S, Lehwark P, Bock R. Organellar Genome DRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 2019;47:59–64.
74. Beier S, Thiel T, Munch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinform. 2017;33:2583–5.
75. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 2004;32:273–9.
76. Amiryousefi A, Hyvönen J, Poczai P, IRscope. An online program to visualize the junction sites of chloroplast genomes. Bioinform. 2018;34:3030–1.
77. Sharp PM, Li WH. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987;15:1281–95.
78. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160–6.
79. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinform. 2009;25:1451–2.
80. Hall TA, BioEdit. A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symp. ser. 1999;41:95–98.
81. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinform. 2014;30:1312–3.
82. Miller MA, Pfeiffer W, Schwartz T. November. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 2010;pp. 1–8.
83. Bouckaert R, Heled J, Denise K, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ, et al. BEAST2: a software platform for bayesian evolutionary analysis. PLoS Comput Biol. 2014;10:e1003537.
84. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018;67:901–4.

Publisher's Note