Clinical Complementary Medicine and Pharmacology 1 (2021) 100002

Contents lists available at ScienceDirect

Clinical Complementary Medicine and Pharmacology

journal homepage: www.elsevier.com/locate/ccmp

Full Length Article

The Complete Chloroplast Genome Sequences of *Anisodus Acutangulus* and a Comparison with Other Solanaceae Species

Qikai Huang[1], Zhixiang Liu[1], Can Wang, Mingyi Jing, Junqiu Liu, Wei Zhou, Guoyin Kai*

Laboratory of Medicinal Plant Biotechnology, College of pharmacy, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, 310053, PR China

ARTICLE INFO

Keywords:
Anisodus acutangulus
Chloroplast genome
Comparative genomic
Phylogenetic analysis
Species identification

ABSTRACT

Anisodus acutangulus (Solanaceae), an important folk medicinal herb in China, produces up to 1.2% alkaloids more than that in other Solanaceae plants such as *Hyoscyamus niger*, while its evolutionary position in Hyoscyameae is not very clear. Objective: To explain the evolutionary position of *A. acutangulus* in the Solanaceae via complete chloroplast genome(cp) sequence. Methods: Complete chloroplast genome of *A. acutangulus* was obtained and characterized using the Illumina PE150 pair-end sequencing data. Structure of the genome, codon usage, nucleotide variability (Pi) value, distribution of repeats and SSRs between *A. acutangulus* and other seven Solanaceae species were analyzed. Previously published 22 Solanaceae cp genomes were used to construct phylogenetic tree. Results: The complete cp genome of *A. acutangulus* is 156082 bp in length, showed the typical quadripartite structure. The complete cp genome of *A. acutangulus* was highly conserved. A total of 112 unique genes were found in cp genome of *A. acutangulus,* among which 17 were duplicated. Further, we found eight hotspot regions for genome divergence could be explored as new DNA barcodes for the identification of the Solanaceae species. Phylogenetic analysis showed that *A. acutangulus* formed a clade with *H. niger.* Conclusion: *A. acutangulus* belongs to Hyoscyameae subfamily and the complete cp genome provides valuable information for phylogenetic reconstruction or comparative genomics of *A. acutangulus.*

1. Introduction

The nightshade family (Solanaceae) distributes worldwide with about 90 genera and 3000-4000 species, and has economically important nutritive, ornamental, and medicinal value (http://www.Solanaceaesource.org/) (Olmstead and Bohs, 2007, Särkinen et al., 2013, Otálora and Berndt, 2018). Used as common edible fruits and tubers in life, the Solanaceae species include the tomato, potato, eggplant, chilli pepper and so on (William and Zhang, 1992, Vorontsova and Knapp, 2012). For medicinal value, like *Lycium barbarum* and*Solanum nigrum* have been used as traditional Chinese medicines for thousands of years in China, there are a few genera of Solanaceae can produce tropane alkaloids (TAs), such as species of *Hyoscyamus, Datura, Duboisia, Atropa and Scopolia* (Zhang et al., 2004).

Anisodus acutangulus (Hyoscyameae), as an excellent source of TAs, is a perennial and endangered herb of the tribe Hyoscyameae (Solanaceae) endemic to Yunnan of China (Cui et al., 2015). It has been used as a folk medicine for hundreds of years and is mainly used for the treatment of fracture, rheumatism, lumbago and leg pain, bruise and swelling. As a tribe of Solanaceae, all species of Hyoscyameae are rich in TAs. Usually, plants with the same chemical constituents are more closely related in the modern chemotaxonomical systems (Martins and Nunez, 2015, Pigatto et al., 2015). A new subfamily Atropoideae was established in 1987 (Tétény, 1987). This subfamily is characterized by the production of TAs, according to the characteristics of external morphology, palynology and phytochemical composition (Hoare and Knapp, 1997). However, classic taxonomy and chemotaxonomical methods have their limitations, modern (DNA-based) molecular plant systematics which is more properly for phylogenetic analysis often shows different results (Olmstead et al., 1999, Volis et al., 2018, Tu et al., 2010, Gates et al., 2018), such as the evolutionary position in Solanaceae of *Atropa* and *Mandragora*. So, the relationships among the taxa of the Hyoscyameae remain unclear.

Chloroplast, as the organelle of photosynthesis, is the most important and common plasmid in plant cells. Its own genome is conserved throughout higher plants at the structural and genic level (Cho et al., 2015, Daniell et al., 2016). The cp genome is an exposed circular double-stranded DNA molecule of about 120-210 kb (Palmer, 1985). For most plants, the cp genome is characterized by two inverted repeat (IRA and IRB) regions, a large single-copy (LSC) region, and a small single-

* Corresponding author.
E-mail addresses: 979845003@qq.com (Q. Huang), zhixiangliu88@163.com (Z. Liu), 2281010784@qq.com (C. Wang), jingmingyi2018@163.com (M. Jing), junqiuliu@zcmu.edu.cn (J. Liu), zhouwei19810501@163.com (W. Zhou), guoyinkai@yahoo.com, kaiguoyin@zcmu.edu.cn (G. Kai).

[1] These authors contributed equally.

https://doi.org/10.1016/j.ccmp.2021.100002
Received 26 April 2021; Received in revised form 24 May 2021; Accepted 27 May 2021

Table 1
Genes in the chloroplast genomes of *A. acutangulus*.

Gene Category	Gene Groupes	Gene Names
Transcription and translation	Small subunit of ribosome	rps12[2,3], rps16[1], rps2, rps3, rps4, rps7[3], rps11, rps8, rps18, rps14, rps19, rps15
	Large subunit of ribosome	rpl2[1,3], rpl16[1], rpl22, rpl20, rpl14, rpl23[3], rpl33, rpl32, rpl36
	rRNA genes	rrn5[3], rrn16[3], rrn4.5[3], rrn23[3],
	tRNA genes	trnfM-CAU, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnI-AUG[3], trnL-CAA[3], trnL-UAG, trnM-CAU, trnN-GUU[3], trnP-UGG, trnQ-UUG, trnR-ACG[3], trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC[3], trnW-CCA, trnY-GUA, trnA-UGC[1,3], trnG-UCC[1], trnI-GAU[1,3], trnK-UUU[1], trnL-UAA[1], trnV-UAC[1]
	DNA-dependent RNA polymerase	rpoB, rpoA, rpoC2, rpoC1[1]
Genes for photosynthesis	Subunits of photosystem I	psaA, psaB, psaC, psaI, psaJ, ycf3[2], ycf4
	Subunits of photosystem II	psbB, psbC, psbA, psbD, psbE, psbH, psbZ, psbK, psbN, psbJ, psbF, psbM, psbT
	Subunits of cytochrome	petA, petB[1], petD[1], petG, petL, petN
	Subunits of ATP synthase	atpA, atpB, atpE, atpF[1], atpH, atpI
	Large subunit of rubisco	rbcL
	NADH oxidoreductase	ndhA[1], ndhB[1,3], ndhF, ndhD, ndhH, ndhK, ndhG, ndhI, ndhJ, ndhC, ndhE
Other genes	other	accD, ccsA, cemA, clpP[2], matK
Unknown	function protein-coding gene	ycf1[3], ycf2[3]

[1] Gene containing a single intron
[2] Gene containing two introns
[3] shows genes duplicated.

copy (SSC) region. Since the cp genome sequence can easy to obtain and its size and nucleotide substitution rate are moderate, it has been widely used to analysis plant phylogenies (Clegg et al., 1994). With the rapid development of sequencing technology, more and more cp genomes have been sequenced and reported, and the application of phylogenetic analysis with complete cp genome has been growing annually (Song et al., 2017, Gu et al., 2019, Xue et al., 2019, Liu et al., 2018, Kim et al., 2019, Lee et al., 2019, Park et al., 2018).

In order to clarify the evolutionary position of *A. acutangulus* in Hyoscyameae, we got a new cp genome sequenced by Illumina HiSeq 4000 Platform, and reconstructed a new molecular phylogeny using the cp genome sequences of Hyoscyameae. In this study, we analyzed the structure of cp genome, codon usage, distribution of repeats, and SSRs by comparing with previously published cp genome of various genera species in Hyoscyameae which also can generate TAs. Finally, based on a total of 22 complete cp genomes of Solanaceae, the new phylogenetic relationships were estimated. Our study will provide the complete cp sequence data of *A. acutangulus*, and the comparative phylogenetic and molecular evolutionary analysis of several Solanaceae species rich in TAs, which can be helpful to gene engineering as well as for molecular breeding for these endangered herbal species.

2. Materials and Methods

2.1. Plant material preparation and sequencing

Fresh young leaves of *A. acutangulus* were obtained from asepsis seedlings being cultivated in the plant growth chamber Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China. Total genomic DNA was extracted with modified 2 × cetyltrimethyl ammonium bromide (CTAB) DNA-extraction method (Doyle and Doyle, 1986). The extracted DNA was sheared into 300-400 bp fragments with a Covaris M220 (Covaris, United States) and built a shotgun library following the procedure of NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, United States). The library was paired-end sequenced on the Illumina HiSeq 4000 platform. All chloroplast genome sequences used in this study were downloaded from GenBank (Table 1, Table S1).

2.2. Chloroplast genome assembly and annotation

With the complete genome of *A. belladonna* chloroplast (GenBank accession NC_004561) as a reference sequence, we selected *A. acutangulus* chloroplast genome contigs from the Illumina sequencing data adopting the BLAST method. The contigs were assembled using SOPAdenovo2 with default parameters (Luo et al., 2012). Then the scaffolds were used as seed sequences to finish the cp genome sequence by NOVOPlasty (Dierckxsens et al., 2017). Gene annotation of the *A. acutangulus* cp genome was performed using the web application GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq.html) (Tillich et al., 2017). The circular cp genome map of the *A. acutangulus* was drawn by OGDRAW (http://ogdraw.mpimpgo-lm.mpg.de/) (Lohse et al., 2013) and then manually edited by Geneious10.3 (Kearse et al., 2012).

2.3. Chloroplast genome sequence analyses

The relative synonymous codon usage (RSCU) was analyzed with CodonW1.4.4 (Thompson et al., 2002). Repeat sequences (including forward, reverse, palindromic, and complementary repeats) were analysed using REPuter Online software (https://bibiserv.cebitec.uni-bielefeld.de/reputer/) (Kurtz and Schleiermacher, 1999) with the parameters were set as follows: Hamming distance of 3 and minimum repeat size of 30 bp. Simple sequence repeats (SSRs) were detected by MISA (https://webblast.ipk-gatersleben.de/misa/). Thresholds for a minimum number of repeat units were established as follows: > 10 for mono-nucleotide, > 5 for di-nucleotide, > 4 for tri-nucleotide, and > 3 for tetra-nucleotide, penta-nucleotide and hexa-nucleotide SSR. The cp genomes of the eight Solanaceae species were aligned with MAFFT, visualized using mVISTA (http://genome.lbl.gov/vista/mvista/submit.shtml) (Frazer et al., 2004) in Shuffle-LAGAN mode, with *A. belladonna* cp genome annotation as a reference. DnaSP v5 (Librado and Rozas, 2009) was used to analyze the nucleotide diversity (Pi) among the cp genomes of the eight species, basing on the sliding window analyses. The window length was 600 bp and step size was 200 bp.

2.4. Phylogenetic analyses

We selected 22 complete cp genomes obtained from GenBank (Table S1), including 21 Solanaceae species and 1 Scrophulariaceae species defined as the outgroups for phylogenetic trees analyses. The Maximum Parsimony (MP) analysis in PAUP4.0 (Cummings, 2004) was used to construct MP tree, while the Maximum Likelihood (ML) tree was made by using RAxML (Stamatakis, 2006) with general Time-Reversible, gamma distribution (GTR + G) model and 1000 bootstrap replicates. For Bayesian Inference (BI) analysis, MrBayes (Huelsenbeck and Ronquist, 2001) was used with Markov chain Monte Carlo algorithm, running 2000000 generations with trees sampled every 1000 generations

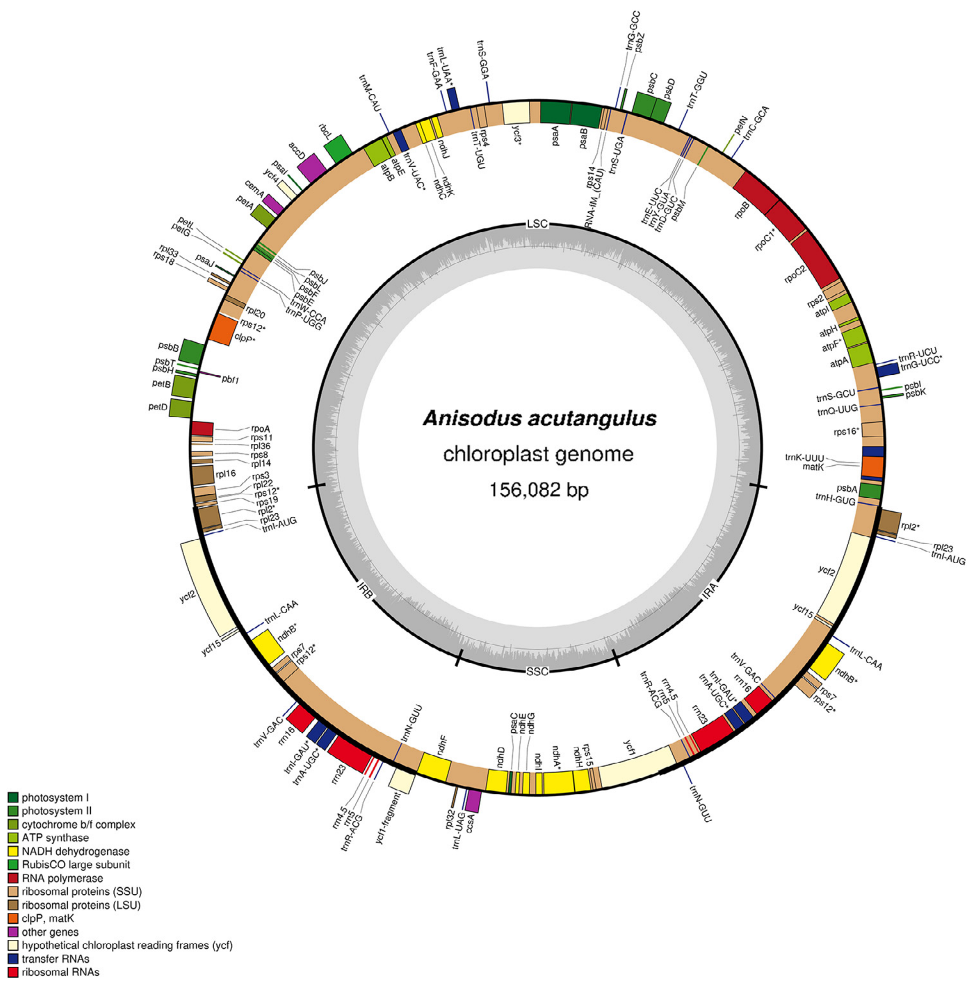

Fig. 1. Chloroplast genome map of *A. acutangulus*. Genes on the outside of the ring represent counterclockwise transcription and genes on the inside of the ring clockwise transcription. The thick black lines in the outer ring represent two IR regions. The dark gray graph of the kernel represents the GC content. Small single copy (SSC), large single copy (LSC) and inverted repeats (IRa, IRb).

and the 250000 samples discard of the trees. When the average standard deviation is less than 0.01, it means that the stationarity was reached.

3. Results and Discussion

3.1. Anisodus acutangulus *chloroplast genome*

Based on 2.01G paird-end Illumina sequencing reads of *A. acutangulus*, we yielded a new complete circular chloroplast genome of *A. acutangulus* with a quadripartite structure sequences of 156,082 bp in length. It was consisted of a pair of the inverted repeats (IRs), LSC and SSC regions with 25906 bp, 86530 bp and 17741 bp in length (Fig. 1), respectively. The GC content of *A. acutangulus* was 37.6%, and IR regions had higher GC contents (42.9%) than LSC regions (35.6%) and SSC regions (31.9%).

A total of 113 unique genes were found in cp of *A. acutangulus*, including 78 protein-coding genes, 2 conserved hypothetical chloroplast reading frames (*ycfs*), 30 transfer RNA genes (tRNA) and 4 ribosomal RNA genes (rRNA), not counting identical copies (Table 1). 18 genes were duplicated in the IR, including seven protein-coding genes (*rps12, rps7, rpl2, rpl23, ndhB, ycf1, ycf2*), seven tRNA (*trnI-AUG, trnL-CAA, trnN-GUU, trnR-ACG, trnV-GAC, trnA-UGC,* and *trnI-GAU*) and four rRNA (*rrn5, rrn16, rrn4.5,* and *rrn23*).

Twelve of the protein-coding genes contained introns, of which, nine (*rps16, rpl2, rpl16, rpoC1, petB, petD, atpF, ndhA,* and *ndhB*) had one intron and three (*rps12, ycf3,* and *clpP*) contained two introns. Except for intron 1 in *rps12* and the *trnL-UAA* intron that are trans-spliced, the rest are cis-spliced introns. Six rRNA genes (*trnA-UGC, trnG-UCC, trnI-GAU, trnK-UUU, trnL-UAA,* and *trnV-UAC*) contained one intron. But there were some exceptions that non-ATG codons were identified as start codons, such as GUG in *rps19* and *ndhD*. It was a common feature in land plants for a variety of chloroplast genes to use ACG or GUG rather than the canonical AUG as start codon (Hirose et al., 1999, Raubeson et al., 2007).

3.2. *Comparative chloroplast genomic analysis*

Changes in chloroplast genome sizes are mostly as a result of the expansion and contraction of the border regions. The changes affect the size of cp genomes over a period of time, as a marker for the evolution of chloroplast genomes (Liu et al., 2017). There were seven chloroplast genomes from seven different genera within the Solanaceae have been reported (Table 2). The chloroplast genome of *A. acutangulus* was highly similar to others within the family, with 97.6%, 99,6%, 98.1%, 96.2%, 98.4%, 98.3% and 98.5% identity to *Hyoscyamus niger, Anisodus tanguticus, Atropa belladonna, Datura stramonium, Scopolia parviflora, Physochlaina orientalis, Atropanthe sinensis,* respectively.

Using mVISTA with the annotation of *A. belladonna* cp genome as a reference, the compared result among eight Solanaceae species (Fig. 2) showed that most regions were conserved, especially the IR region, related to the fact that IR regions are more conserved in evolution. The coding regions were more conserved than the non-coding regions. *ndhF, ycf1* In the coding regions, and the non-coding region located in 4-8 k, 27-34 k, 44-50 k, and 65-70 k had highly divergent among eight Solanaceae species. According to the nucleotide variability (Pi) (Fig. 3), the IR regions were more conserved than single-copy regions. We found eight hotspot regions for genome divergence could be new

Table 2
The basic characteristics of chloroplast genomes of eight Solanaceae species.

Species	A.acutangulus	H. niger	A.tanguticus	Atropa belladonna	Datura stramonium	Scopolia parviflora	Physochlaina orientalis	Atropanthe sinensis
Accession number	MT558919	KF248009	MK347419	NC_004561	NC_018117	NC_030282	NC_044154	NC_044471
Total cp genome size (bp)	156082	155720	155767	156687	155871	156193	156371	156565
LSC region (bp)	86530	86105	86515	86869	86299	86364	86598	86600
IR region (bp)	25906	25876	25881	25901	25602	25905	25861	25939
SSC region (bp)	17741	17864	17487	18008	18366	18019	17989	18087
Total number of genes (unique)	112	112	112	112	113	112	112	112
Protein-coding gene (unique)	78	78	78	78	79	78	78	78
rRNA (unique)	4	4	4	4	4	4	4	4
tRNA (unique)	30	30	30	30	30	30	30	30
GC content (%)	37.6	37.6	37.6	37.6	37.9	37.6	37.7	37.6
GC content of LSC (%)	35.6	35.6	35.6	35.6	36.0	35.7	35.8	35.7
GC content of IR (%)	42.9	42.9	42.9	42.9	43.1	42.9	42.9	42.9
GC content of SSC (%)	31.9	31.5	31.9	31.7	32.3	31.8	32.0	31.9

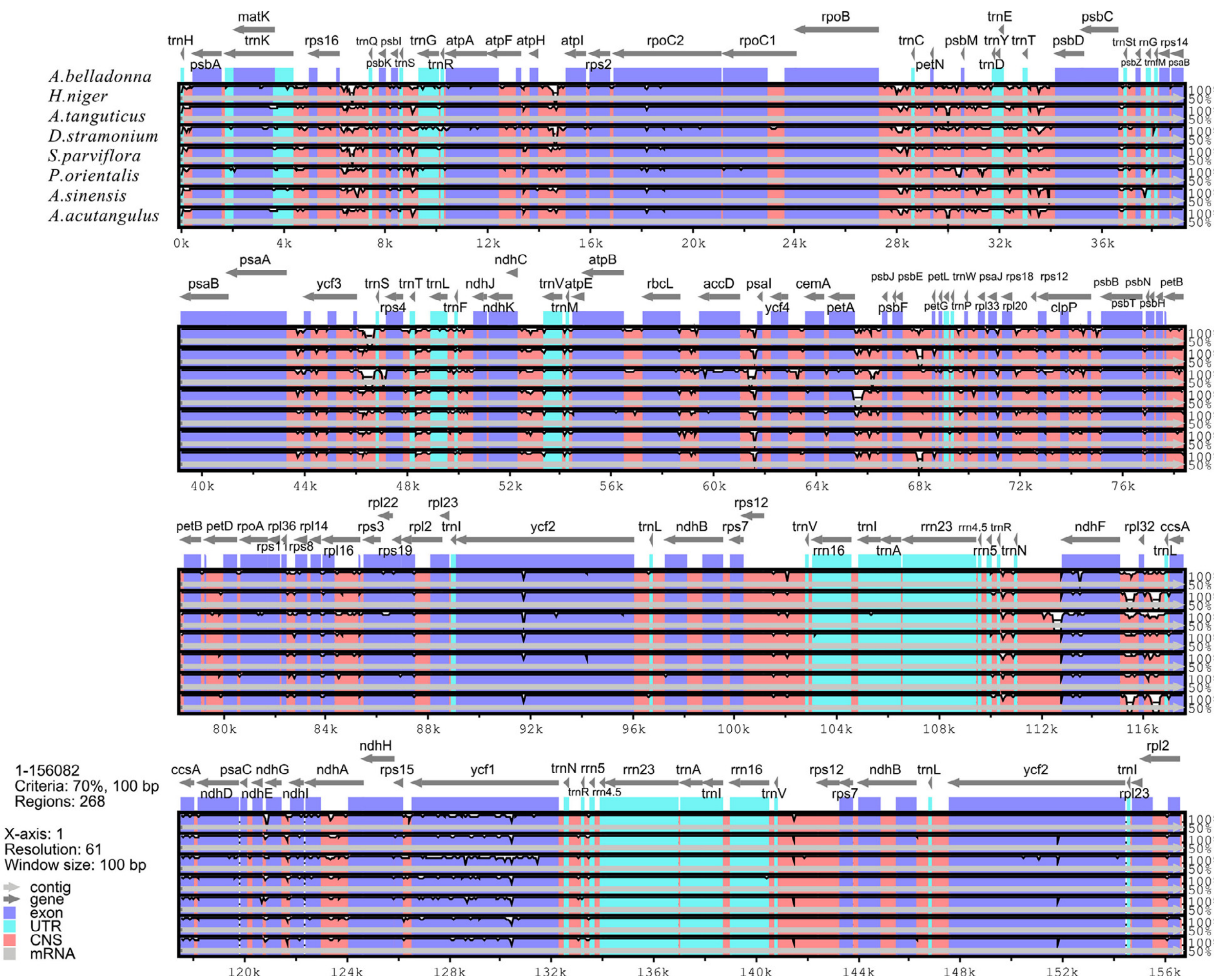

Fig. 2. Codon content in the *A.acutangulus*, RSCU: relative synonymous codon usage.

DNA barcodes for species identification (Xue et al., 2019, Dong et al., 2017). Those regions were *trnH-psbA, trnK-rps16, rps16-trnQ, rpoB-trnC, rpl36-rps8, ndhF-rpl32, rpl32-trnL*, and *ycf1*. The contraction and expansion of IR borders can reflect the phylogenetic relationship of species (Zhang et al., 2017). The structure variation could be found in IRs/SC borders between eight species (Fig. 4). In the four species (*H. niger, A. acutangulus, D. stramonium, P. orientalis*), the *ndhF* gene overlapped with the *ycf1*. Comparied with others, *ndhF* gene in *A. acutangulus* and *D.stramonium* was closer to IRb. The *rps19* gene located in LSC/IRa border, and in IRa region of *A. acutangulus* and *A. tanguticus,* has the same length (75 bp). As well, the *trnH* gene separated from the IRA/LSC border by a spacer varies from 14 bp. The *ycf1* gene spanned the SSC/IRa region, and *ycf1* gene in *D. stramonium* has more parts in SSC region.

3.3. Codon usage

Codon usage bias also called Relative synonymous codon usage (RSCU) is the variation in the frequency of occurrence of synonymous codons in coding DNA. As an essential evolutionary feature, it is of great significance to master the codon usage bias in different species

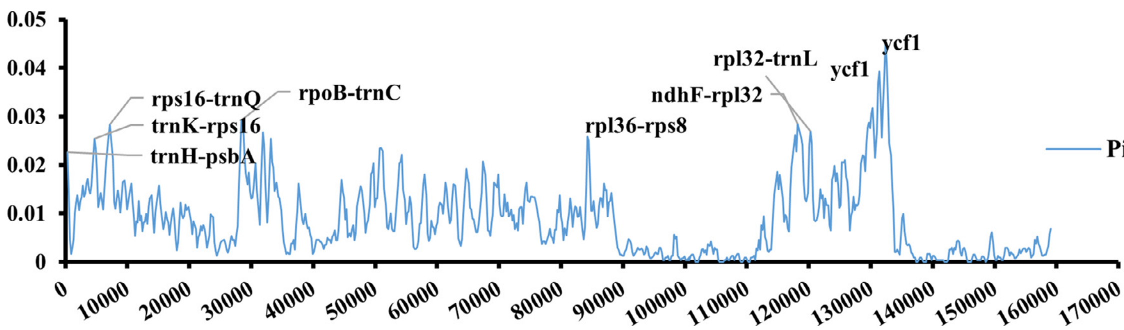

Fig. 3. A: Number of different repeat types; B: Number of different repeat lengths

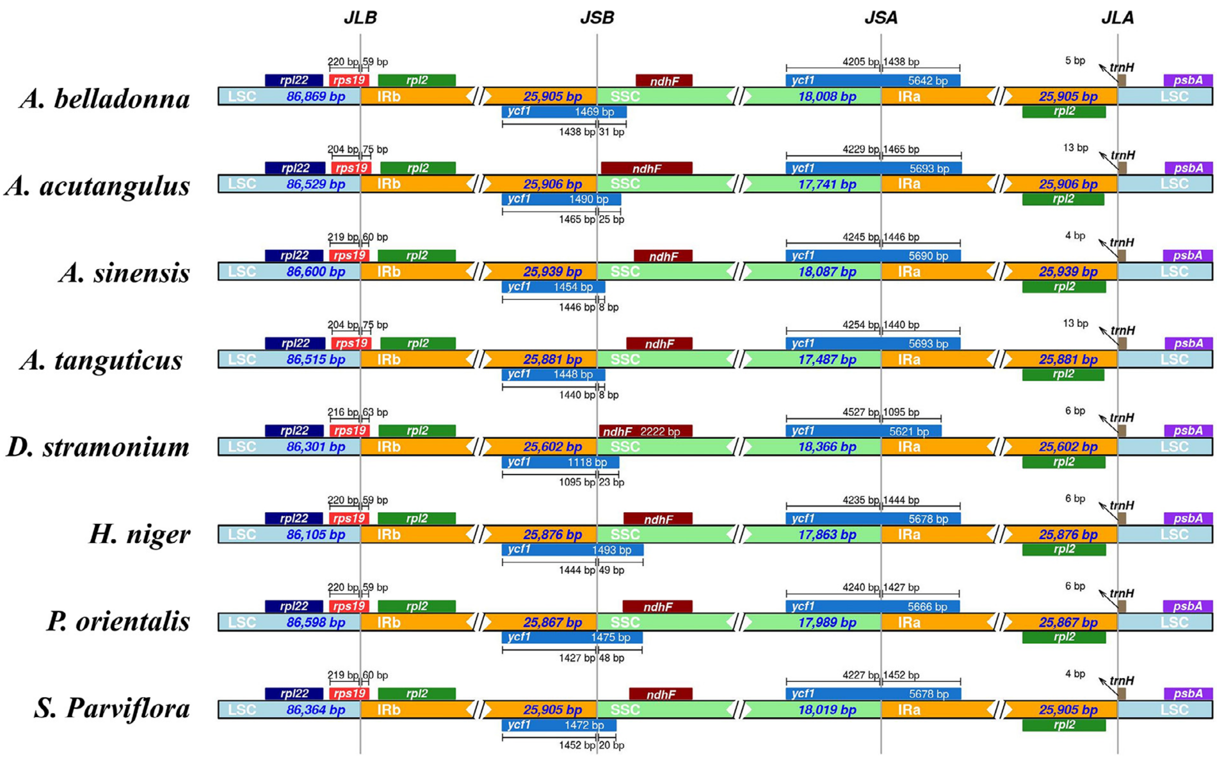

Fig. 4. The number and distribution of SSRs in the chloroplast genomes of eight Solanaceae species. A: Proportion of repeats in LSC, SSC, IR regions; B: Number of repeats in LSC, SSC and IR; C: Number of different repeat types.

(Yan et al., 2019, Feng et al., 2013). Here, we analyzed RSCU value to learn the codon usage of *A. acutangulus* cp genomes, in which RSCU > 1 represents the preference of the codon and RSCU < 1 indicates the low usage of the codon. The protein-coding region of the *A. acutangulus* cp genomes was encoded by 26,900 codons (Fig. 5 and Table S2), most of the preferred amino acid-encoding codons had A or U as the third nucleotide. This phenomenon has been found in other species (Park et al., 2017). By contraries, C or G as the third in amino acid-encoding codons had RSCU < 1. The most and least universal amino acids of *A. acutangulus* cp genomes are leucine (10.6%) and cysteine (1.1%), respectively. The most codon was AUU with a total of 1110, encoding isoleucine, while the least codons were UGC with 68, encoding cysteine. The AUG and UGG, which encoding methionine and tryptophan, showed no bias (RSCU = 1). Furthermore, the codon usage of *A. acutangulus* did not show much difference compared with other Solanaceae plants (Table S2).

3.4. Repeat and SSR analyses

Repeat sequences provide important information about genomes. Using REPuter, we found some forward, palindromic and reverse repeats in eight species cpDNAs. The number of three types of repeats in *A. acutangulus* cp genome were 21, 21, and 7, respectively (Fig. 6A). The length of repeats ranged from 21 to 48 bp (Fig. 6B), Most repeat sequences with 20-30 bp distributed in the intron and intergenic regions, whereas some were found in genes such as *ycf1, ycf2, ycf3, psaB, and pasA* (Table S3). There were no repeats longer than 50 bp in *A. acutangulus*, but this is not the case in *A. tanguticus, D. stramonium, S. parviflora, H. niger*, and *P. orientalis* having repeats longer than 50 bp. Among these, only four repeats longer than 60 bp can be found in *P. orientalis*.

Simple sequence repeats (SSRs), also known as microsatellites, are 1 to 6 bp repeating sequences extensively distributed in the chloroplast genome. SSRs are highly polymorphic and codominant, which are valu-

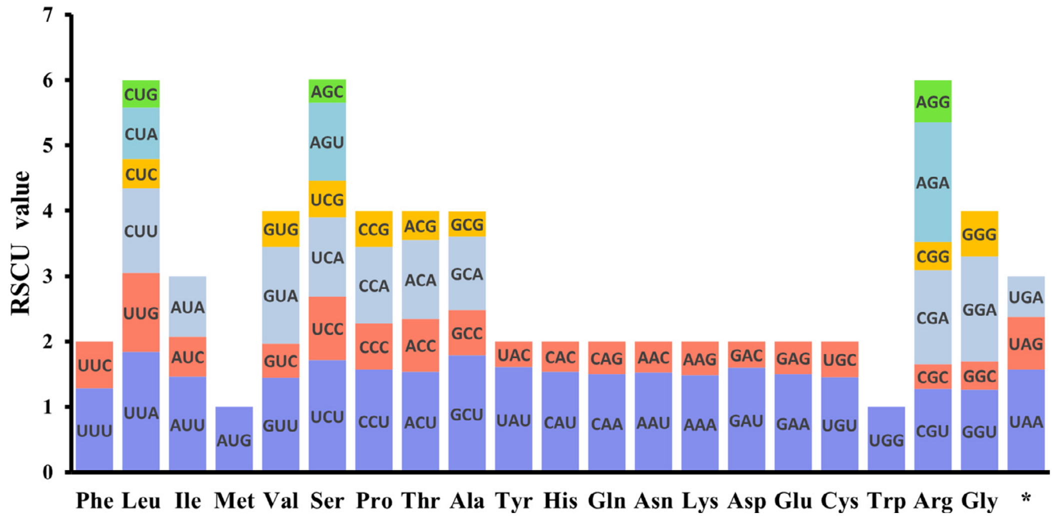

Fig. 5. Comparison of eight chloroplast genomes using *A. belladonna* annotation as a reference. The y-axis represents the percent identity within 50–100%.

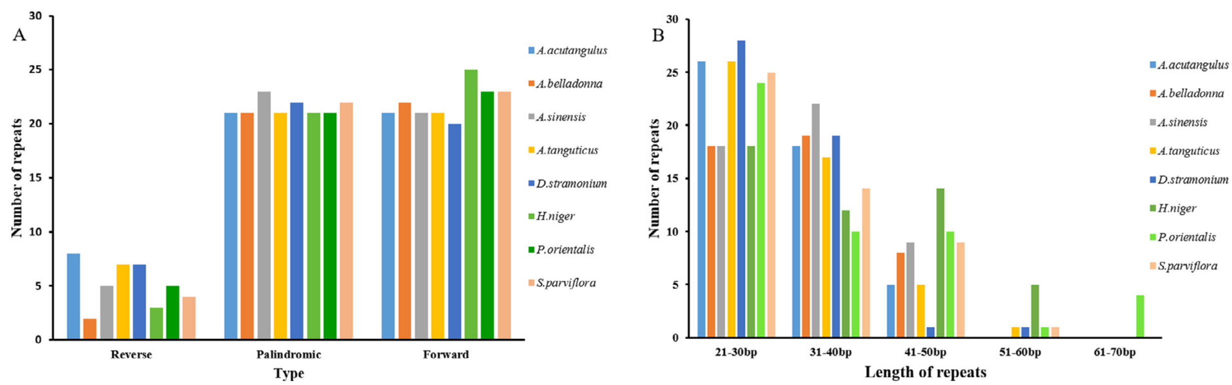

Fig. 6. Nucleotide diversity (Pi) in the complete cp genome of eight Solanaceae species. Sliding window analysis with a window length of 600 bp and a step size of 200 bp.

able markers for a study involving gene flow, population genetics, and gene mapping (He et al., 2012). There were a total of 412 SSRs in eight cp genomes and 53 SSRs in *A. acutangulus*. The number of SSRs ranged from 47 to 59 in eight species, and mono-nucleotides account for most of them. Most of the SSR repeats located in LSC (76.2%), while lest of the SSRs situated in SSC (10.6%) (Fig. 7A). All of *A. acutangulus* and other seven Solanaceae species have a similar ratio for SSRs (Fig. 7B and Table S4). Compared to other Solanaceae species, *P. orientalis* had the highest number of SSRs with 59, while *A. tanguticus* had the least (Fig. 7C). The SSRs of *A. acutangulus* were composed of 37 mononucleotides, 8 dinucleotides, 1 trinucleotide, 6 tetranucleotides, and 1 pentanucleotide. The mono-nucleotide SSRs are A/T, and di-nucleotide SSRs are AT/TA, enriching AT content of the cp genomes. The cp genome of *A. sinensis* has a TTTATA hexa-nucleotide SSRs, but none in others. Those identified repeats would help population genetics and phylogenetic studies in Solanaceae.

3.5. Phylogenetic analyses

With *Pedicularis ishidoyana* as the outgroup, the phylogenetic relationship of eight Solanaceae was analyzed by maximum parsimony (MP), maximum likelihood (ML) and Bayesian analysis (BI), respectively. The three methods showed the similar topologies (Fig. 8, S1, S2). *A. acutangulus* with *A. tanguticus* and *H. niger* formed one branch. In addition, *A. belladonna* was sister to the rest of the genera of the Hyoscyameae, which was consistent with the previous report (Olmstead et al., 2008). Tropane alkaloids and calystegines existed and distributed in Solanaceae reported in previous studies (El Bazaoui et al., 2011, Alvarenga et al., 2001, Wink, 2003, Doncheva et al., 2006). According to plant chemotaxonomy, it was suggested that *Lycium* was closely related to *Capsicum*, and *Datura* is sister to a clade containing *Anisodus* (Pigatto et al., 2015). Besides, most of the secondary metabolites of *D. stramonium* were identical to Hyoscyameae species but different to *Lycium* species. However, this study showed that the *Lycium* displayed closer relationship with the Hyoscyameae than *D. stramonium*. Direct analysis by real time-high resolution mass spectrometry revealed that *Atropa* and *Datura* form a clade (Beyramysoltan et al., 2019), but it was found that the relationship between *Datura* and *Atropa* is far away in our study. It was reported that *D. stramonium* was classified into the Datureae according to molecular plant systematics studies (Jamil et al., 2014), but our result showed *D. stramonium* was categorized into the Solaneae. Comparing to phylogeny relationship using individual genes such as *ITS*, *rbcl*, *ndhF* or *trnL*, it is more accurate to establish based on complete cp genomes. Our results strongly supported the new classification system of the Hyoscyameae, and clarified the evolutionary position of *A. acutangulus* in Hyoscyameae.

4. Conclusion

The cp genomes of *A. acutangulus* were sequenced and annotated, and the sequencing data was one of valuable resources for evolutionary relationships among Solanaceae. By comparing with other Solanaceae species, we found the structure and composition of *A. acutangulus* cp genomes are in a high degree of similarity with other Hyoscyameae

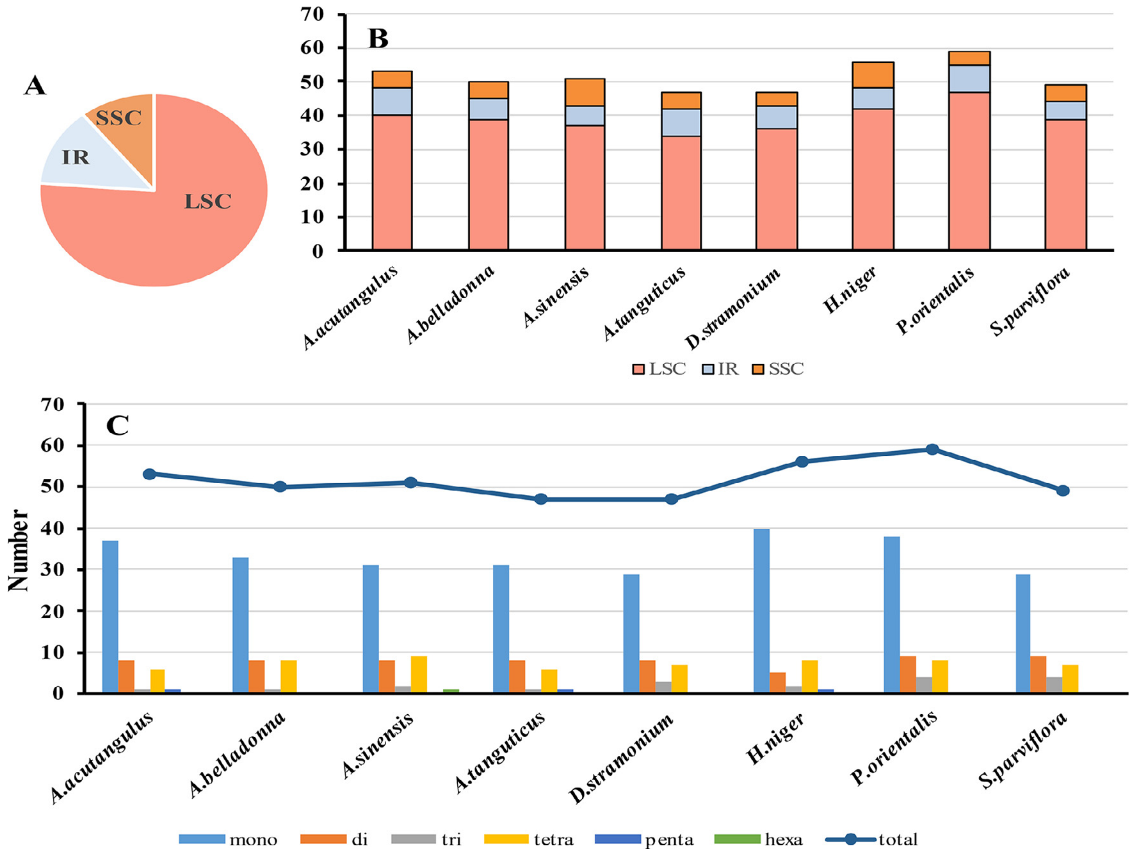

Fig. 7. Chloroplast genome borders in eight Solanaceae species. LSC (large single copy region), SSC (small single copy region), and IR (inverted repeat region).

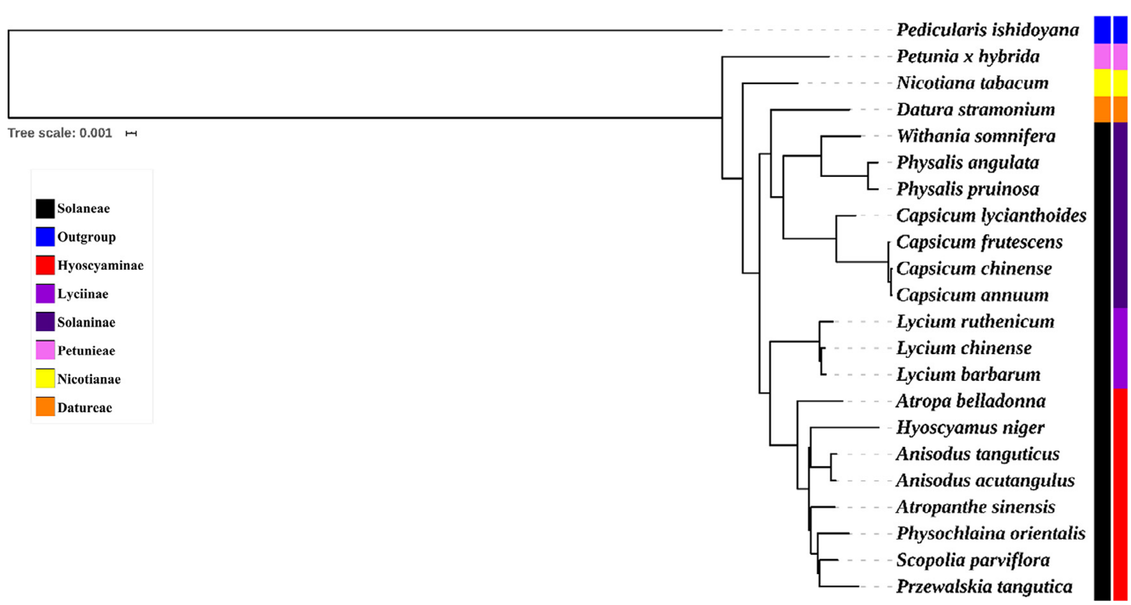

Fig. 8. Maximum Likelihood tree based on the complete chloroplast genome.

species. Maybe the phylogenetic relationship between *D. stramonium* and *A. acutangulus* is relatively far away, so it's hard to find out *psbL* protein-coding gene in *A. acutangulus* and other Hyoscyameae species apart from *D. stramonium.* Eight hotspot regions (*trnH-psbA, trnK-rps16, rps16-trnQ, rpoB-trnC, rpl36-rps8, ndhF-rpl32, rpl32-trnL*, and *ycf1)* were found, which could be used as new DNA barcodes for species identification. The complete cp genomes of seven species from Hyoscyameae, and one from Datureae was focused on the structural and gene comparison. A total of 22 complete cp genomes from Solanaceae were used for phylogenetic reconstruction, It showed that *A. acutangulus* was close to Hyoscyameae (because of that *A. tanguticus* and *H. niger* formed one branch) and *D. stramonium* was close to Solaneae. Those results may be beneficial to the classification and phylogeny reconstruction of *A. acutangulus*.

Data Availability

Chloroplast genome sequence of *A. acutangulus* can be accessed via accession number MT558919 in NCBI GenBank.

Funding

This work was financially supported by the National Key R&D Program of China (2018YFC1706200), National Natural Science Fund of China (81522049, 31571735, 82003888), Zhejiang Provincial Ten Thousands Program for Leading Talents of Science and Technology Innovation (2018R520), Zhejiang Provincial Program for the Cultivation of High-level Innovative Health talents.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Qikai Huang: Formal analysis, Software, Writing – original draft. **Zhixiang Liu:** Methodology, Writing – review & editing. **Can Wang:** Writing – review & editing. **Mingyi Jing:** Visualization. **Junqiu Liu:** Resources. **Wei Zhou:** Supervision. **Guoyin Kai:** Conceptualization, Validation, Funding acquisition.

Acknowledgements

We thank Guoyin Kai and Zhixiang Liu for conceiving and designing the experiments, and we thank all the colleagues in this study for providing useful discussions and technical assistance.

ORCID

Guoyin Kai, https://0000-0001-7586-906.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ccmp.2021.100002.

References

Alvarenga, S.A.V., Ferreira, M.J.P., Emerenciano, V.P., Cabrol-Bass, D., 2001. Chemosystematic studies of natural compounds isolated from Asteraceae: characterization of tribes by principal component analysis. Chemometrics Intell. Lab. Syst. 56 (1). doi:10.1016/S0169-7439(01)00103-4.

Beyramysoltan, S., Abdul-Rahman, N.H., Musah, R.A., 2019. Call it a "nightshade"-A hierarchical classification approach to identification of hallucinogenic Solanaceae spp. using DART-HRMS-derived chemical signatures. Talanta 204, 739–746. doi:10.1016/j.talanta.2019.06.010.

Cho, K.S., Yun, B.K., Yoon, Y.H., et al., 2015. Complete chloroplast genome sequence of Tartary Buckwheat (Fagopyrum tataricum) and comparative analysis with common Buckwheat (F. esculentum). PLoS One 10 (5), e0125332. doi:10.1371/journal.pone.0125332.

Clegg, M.T., Gaut, B.S., Learn Jr., G.H., Morton, B.R., 1994. Rates and patterns of chloroplast DNA evolution. Proc. Natl. Acad. Sci. U. S. A. 91 (15), 6795–6801. doi:10.1073/pnas.91.15.6795.

Cui, L., Huang, F., Zhang, D., et al., 2015. Transcriptome exploration for further understanding of the tropane alkaloids biosynthesis in *Anisodus acutangulus*. Mol. Genet. Genomics 290 (4), 1367–1377. doi:10.1007/s00438-015-1005-y.

Cummings, M.P., 2004. PAUP* [Phylogenetic Analysis Using Parsimony (and Other Methods)]. In Dictionary Bioinf. Comput. Biol. doi:10.1002/0471650129.dob0522.

Daniell, H., Lin, C.S., Yu, M., Chang, W.J., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. 17 (1), 134. doi:10.1186/s13059-016-1004-2.

Dierckxsens, N., Mardulyn, P., Smits, G., 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic. Acids. Res. 45 (4), e18. doi:10.1093/nar/gkw955.

Doncheva, T., Berkov, S., Philipov, S., 2006. Comparative study of the alkaloids in tribe Datureae and their chemosystematic significance. Biochem. Syst. Ecol. 34 (6), 478–488. doi:10.1016/j.bse.2006.01.008.

Dong, W., Xu, C., Li, W., et al., 2017. Phylogenetic resolution in Juglans based on complete chloroplast genomes and nuclear DNA sequences. Front. Plant Sci. 8, 1148. doi:10.3389/fpls.2017.01148.

Doyle, J.J., Doyle, J.L., 1986. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry 19, 11–15. doi:10.1016/j.bse.2009.07.003.

El Bazaoui, A., Bellimam, M.A., Soulaymani, A., 2011. Nine new tropane alkaloids from *Datura stramonium* L. identified by GC/MS. Fitoterapia 82 (2), 193–197. doi:10.1016/j.fitote.2010.09.010.

Feng, C., Xu, C.J., Liu, W.L., et al., 2013. Codon usage patterns in Chinese bayberry (*Myrica rubra*) based on RNA-Seq data. BMC Genom. 14, 732. doi:10.1186/1471-2164-14-732.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., Dubchak, I., 2004. VISTA: computational tools for comparative genomics. Nucleic. Acids. Res. 32 (Web Server issue), W273–W279. doi:10.1093/nar/gkh458.

Gates, D.J., Pilson, D., Smith, S.D., 2018. Filtering of target sequence capture individuals facilitates species tree construction in the plant subtribe Iochrominae (Solanaceae). Mol. Phylogenet Evol. 123, 26–34. doi:10.1016/j.ympev.2018.02.002.

Gu, C., Ma, L., Wu, Z., Chen, K., Wang, Y., 2019. Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales. BMC Plant Biol. 19 (1), 281. doi:10.1186/s12870-019-1870-3.

He, S., Wang, Y., Volis, S., Li, D., Yi, T., 2012. Genetic diversity and population structure: implications for conservation of wild soybean (Glycine soja Sieb. et Zucc) based on nuclear and chloroplast microsatellite variation. Int. J. Mol. Sci. 13 (10), 12608–12628. doi:10.3390/ijms131012608.

Hirose, T., Ideue, T., Wakasugi, T., Sugiura, M., 1999. The chloroplast infA gene with a functional UUG initiation codon. FEBS Lett. 445 (1), 169–172. doi:10.1016/s0014-5793(99)00123-4.

Hoare, A.L, Knapp, S., 1997. A phylogenetic conspectus of the tribe Hyoscyameae (Solanaceae). Bulletin of the Natural History Museum Botany.

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17 (8), 754–755. doi:10.1093/bioinformatics/17.8.754.

Jamil, I., Qamarunnisa, S., Azhar, A., Shinwari, Z.K., Qaiser, M., 2014. Subfamilial relationships within solanaceae as inferred from atpβ-rbcl intergenic spacer. Pak. J. Bot. 46 (2), 585–590.

Kearse, M., Moir, R., Wilson, A., et al., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12), 1647–1649. doi:10.1093/bioinformatics/bts199.

Kim, S.H., Yang, J., Park, J., Yamada, T., Maki, M., Kim, S.C., 2019. Comparison of whole plastome sequences between thermogenic skunk cabbage *Symplocarpus renifolius* and nonthermogenic *S. nipponicus* (Orontioideae; Araceae) in East Asia. Int. J. Mol. Sci. 20 (19), 4678. doi:10.3390/ijms20194678.

Kurtz, S., Schleiermacher, C., 1999. REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics 15 (5), 426–427. doi:10.1093/bioinformatics/15.5.426.

Lee, S.R., Kim, K., Lee, B.Y., Lim, C.E., 2019. Complete chloroplast genomes of all six Hosta species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. BMC Genom. 20 (1), 833. doi:10.1186/s12864-019-6215-y.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25 (11), 1451–1452. doi:10.1093/bioinformatics/btp187.

Liu, L.X., Li, R., Worth, J.R.P., et al., 2017. The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myricaceae): implications for understanding the evolution of Fagales. Front. Plant. Sci. 8, 968. doi:10.3389/fpls.2017.00968.

Liu, L., Wang, Y., He, P., et al., 2018. Chloroplast genome analyses and genomic resource development for epilithic sister genera Oresitrophe and Mukdenia (Saxifragaceae), using genome skimming data. BMC Genom. 19 (1), 235. doi:10.1186/s12864-018-4633-x.

Lohse, M., Drechsel, O., Kahlau, S., Bock, R., 2013. OrganellarGenomeDRAW–a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic. Acids. Res. 41 (Web Server issue), W575–W581. doi:10.1093/nar/gkt289.

Luo, R., Liu, B., Xie, Y., et al., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1 (1), 18. doi:10.1186/2047-217X-1-18.

Martins, D., Nunez, C.V., 2015. Secondary metabolites from Rubiaceae species. Molecules. 20 (7), 13422–13495. doi:10.3390/molecules200713422.

Olmstead, R.G., Bohs, L., 2007. A summary of molecular systematic research in solanaceae: 1982-2006. Acta. Hortic. 745, 255–268. doi:10.17660/ActaHortic.2007.745.11.

Olmstead, R.G, Sweere, J.A, Spangler, R.E., Bohs, L., Palmer, J., 1999. Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA.

Olmstead, R.G., Bohs, L., Migid, H.A., Santiago-Valentin, E., Garcia, V.F., Collier, S.M., 2008. A molecular phylogeny of the Solanaceae. Taxon 57, 1159–1181. doi:10.1002/tax.574010.

Otálora, M.A.G., Berndt, R., 2018. A taxonomic revision of the genus Puccinia on Lycieae, a tribe of Solanaceae. Mycologia 110 (4), 692–709. doi:10.1080/00275514.2018.1478538.

Palmer, J.D., 1985. Comparative organization of chloroplast genomes. Annu. Rev. Genet. 19, 325–354. doi:10.1146/annurev.ge.19.120185.001545.

Park, I., Kim, W.J., Yeo, S.M., et al., 2017. The complete chloroplast genome sequences of *Fritillaria ussuriensis* Maxim. and *Fritillaria cirrhosa* D. Don, and comparative analysis with other Fritillaria species. Molecules 22 (6), 982. doi:10.3390/molecules22060982.

Park, I., Yang, S., Kim, W.J, Noh, P., Lee, H.O., Moon, B.C., 2018. The complete chloroplast genomes of six *Ipomoea* species and indel marker development for the discrimination of authentic pharbitidis semen (Seeds of *I. nil* or *I. purpurea*). Front. Plant Sci. 9, 965. doi:10.3389/fpls.2018.00965.

Pigatto, A.G., Blanco, C.C., Mentz, L.A., Soares, G.L., 2015. Tropane alkaloids and calystegines as chemotaxonomic markers in the Solanaceae. An. Acad. Bras. Cienc. 87 (4), 2139–2149. doi:10.1590/0001-3765201520140231.

Pigatto, A.G, Blanco, C.C, Mentz, L.A, Soares, G.L., 2015. Tropane alkaloids and calystegines as chemotaxonomic markers in the Solanaceae. An. Acad. Bras. Cienc. 87 (4), 2139–2149. doi:10.1590/0001-3765201520140231.

Raubeson, L.A, Peery, R., Chumley, T.W., et al., 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nuphar advena* and *Ranunculus macranthus*. BMC Genom. 8, 174. doi:10.1186/1471-2164-8-174.

Särkinen, T., Bohs, L., Olmstead, R.G., Knapp, S., 2013. A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol. Biol. 13, 214. doi:10.1186/1471-2148-13-214.

Song, Y., Wang, S., Ding, Y., Xu, J., Li, M.F., Zhu, S., Chen, N., 2017. Chloroplast genomic resource of *Paris* for species discrimination. Sci. Rep. 7 (1), 3427. doi:10.1038/s41598-017-02083-7.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21), 2688–2690. doi:10.1093/bioinformatics/btl446.

Tétény, P., 1987. A chemotaxonomic classification of the Solanaceae. Ann. Mo. Bot. Gard. 74 (3), 600–608. doi:10.2307/2399328.

Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics. 2. doi:10.1002/0471250953.bi0203s00.

Tillich, M., Lehwark, P., Pellizzer, T., et al., 2017. GeSeq-versatile and accurate annotation of organelle genomes. Nucleic. Acids. Res. 45 (W1), W6–W11. doi:10.1093/nar/gkx391.

Tu, T., Volis, S., Dillon, M.O, Sun, H., Wen, J., 2010. Dispersals of Hyoscyameae and Mandragoreae (Solanaceae) from the New World to Eurasia in the early Miocene and their biogeographic diversification within Eurasia. Mol. Phylogenet Evol. 57 (3), 1226–1237. doi:10.1016/j.ympev.2010.09.007.

Volis, S., Fogel, K., Tu, T., Sun, H., Zaretsky, M., 2018. Evolutionary history and biogeography of *Mandragora* L. (Solanaceae). Mol. Phylogenet Evol. 129, 85–95. doi:10.1016/j.ympev.2018.08.015.

Vorontsova, M.S, Knapp, S., 2012. A new species of Solanum (Solanaceae) from South Africa related to the cultivated eggplant. PhytoKeys (8) 1–11. doi:10.3897/phytokeys.8.2462.

William, G.D., Zhang, Z., 1992. Notes on the Solanaceae of China and neighboring areas. Novon 2 (2), 124–128. doi:10.2307/3391672.

Wink, M., 2003. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64 (1), 3–19. doi:10.1016/s0031-9422(03)00300-5.

Xue, S., Shi, T., Luo, W., et al., 2019. Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. Hortic. Res. 6, 89. doi:10.1038/s41438-019-0171-1.

Yan, M., Zhao, X., Zhou, J., Huo, Y., Ding, Y., Yuan, Z., 2019. The complete chloroplast genomes of *Punica granatum* and a comparison with other species in Lythraceae. Int. J. Mol. Sci. 20 (12), 2886. doi:10.3390/ijms20122886.

Zhang, L., Ding, R., Chai, Y., Bonfill, M., Moyano, E., et al., 2004. Engineering tropane biosynthetic pathway in *Hyoscyamus niger* hairy root cultures. Proc. Natl. Acad. Sci. U. S. A. 101 (17), 6786–6791. doi:10.1073/pnas.0401391101.

Zhang, X., Zhou, T., Kanwal, N., Zhao, Y., Bai, G., Zhao, G., 2017. Completion of eight Gynostemma BL. (Cucurbitaceae) chloroplast genomes: characterization, comparative analysis, and phylogenetic relationships. Front. Plant Sci. 8, 1583. doi:10.3389/fpls.2017.01583.