Asian Journal of Agriculture and Biology
https://doi.org/10.35495/ajab.2025.042

Original Article

Complete chloroplast genome of three *Solanum* species (Solanaceae) from China: genome structure, comparative analysis, and phylogenetic relationships

Rebecca Jia Yinn Ng[1], Zhihui Chen[2], Lu Tan[3], Douglas Law[1], Wenbo Liao[2], Yangyang Liu[4]*, Shiou Yih Lee[1]*
[1]Faculty of Health and Life Sciences, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia
[2]State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, 510275 Guangzhou, Guangdong, China
[3]Faculty of Liberal Arts, Shinawatra University, Pathum Thani 12160, Thailand
[4]International Joint Research Center for Quality of Traditional Chinese Medicine, Hainan Branch of the Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, 570311 Haikou, Hainan, China

**Corresponding authors' emails:* yyliu@implad.ac.cn; shiouyih.lee@newinti.edu.my
Received: 29 May 2025 / Accepted: 24 July 2025 / Published Online: 09 August 2025

Abstract

Solanum members are economically important crops worldwide; these species lack molecular information. We assembled and annotated the chloroplast (cp) genomes of three *Solanum* species: *Solanum aculeatissimum*, *Solanum lasiocarpum*, and *Soalnum pittosporifolium*. Genome comparative analyses characterised these species, and a phylogenetic tree was reconstructed using maximum likelihood and approximate Bayesian inference methods on the complete cp genome sequence without IRA. The cp genomes of *S. aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium* displayed a quadripartite circular structure that was 155,821 bp, 155,671 bp, and 155,926 bp in length, respectively. *S. aculeatissimum* and *S. lasiocarpum* cp genomes had 131 unique genes, including 86 protein-coding (CDS), 37 transfer RNA, and eight ribosomal RNA genes. *S. pittosporifolium* had an additional CDS gene, bringing the total to 130. The sequence alignment of 23 *Solanum* species revealed four highly variable regions in the cp genome: *ndhC-trnV*-UAC, *petD*, *rpl33-rpl8*, and *ycf1*, when Pi > 0.03. Based on the complete cp genome sequence of 27 Chinese *Solanum* species, phylogenetic analysis showed a monophyletic relationship. The maximum-likelihood and Bayesian inference methods placed *S. aculeatissimum* and *S. lasiocarpum* in the same clade as *Solanum capsicoides*, while *S. pittosporifolium* was placed with *Solanum dulcamara*, *Solanum japonense*, and *Solanum septemlobum*.

Keywords: Brinjal, Genetic resources, Next-generation sequencing, Plastid genome, Solanales

How to cite this article:
Ng RJY, Chen Z, Tan L, Law D, Liao W, Liu Y and Lee SY Complete chloroplast genome of three *Solanum* species (Solanaceae) from China: genome structure, comparative analysis, and phylogenetic relationships. Asian J. Agric. Biol. 2025: e2025042. DOI: https://doi.org/10.35495/ajab.2025.042

Introduction

Solanaceae is a vast group of flowering plants, consisting of more than 90 genera and approximately 3,000 species (Gebhardt, 2016). The classification of Solanacaeae has undergone substantial revisions over the years, with advancements in molecular techniques playing a crucial role in contributing to a more refined understanding of evolutionary relationships within the family. At present, there are seven subfamilies within Solanaceae, including Cestroideae, Goetzeoideae, Nicotianoideae, Petunioideae, Schizanthoideae, Schwenckieae, and Solanoideae (Olmstead et al., 2008).

Solanum, a prominent genus under Solanoideae, consists of more than 1000 species worldwide (POWO, 2024). As one of the largest genera in the family, members of *Solanum* are often treated as crucial crops that contribute significantly to global agriculture (Tynkevich et al., 2022). Owing to their remarkable range of morphological forms that can adapt to diverse ecological habitats, species of *Solanum* are widely distributed to all parts of the world such as Europe, North and South America, Africa, Australia, and Asia (Musarella, 2020). The economic significance of *Solanum* species also extends to medicinal and ornamental applications. A defining and important feature of *Solanum* is its ability to produce alkaloids, which have been harnessed and used in traditional medicine to treat human diseases such as rheumatism, skin disease, diabetes, and fever (Kaunda and Zhang, 2019). Therefore, the understanding of *Solanum* is not only restricted to advancing botanical knowledge but also to the development of agriculture, horticulture, and sustainable resource utilisation.

Historically, traditional classification methods primarily relied on morphological and anatomical characteristics, such as presence of prickles, type of anthers, and stellate hairs (D'Arcy, 1972). With the introduction of DNA sequencing technologies, it has provided a more accurate representation of the evolutionary relationship within *Solanum*. The first molecular taxonomic study on *Solanum* using sequencing data was conducted by Bohs (2005) using the *ndh*F gene. Despite the limited informative site in the 2,000-bp long sequence, the study was able to identify at least 13 major clades to group the closely related species, of which most are still accepted at present. Despite multilocus analysis being carried out to provide a better resolution to the phylogenetic relationship within *Solanum* (Levin et al., 2006; Weese and Bohs, 2007; Särkinen et al., 2008; Tepe et al., 2016), the outcomes were still incomplete. To improve the findings, Gagnon et al. (2021) reconstructed the phylogenetic tree of *Solanum* based on the complete cp genome and nuclear genome sequences. Their finding has provided a better insight into the molecular classification of *Solanum*, indicating that the genome-based data set is potentially useful in resolving the complex relationship in Solanum at some level. Although a subset of species was sequenced and assembled for their complete cp genome sequences during that time, due to the diversity in *Solanum*, not all species are included in these studies; many *Solanum* species are still understudied.

Next-generation sequencing (NGS) has revolutionised genomics research by facilitating rapid and high-throughput sequencing of complete genomes. In plant phylogenetics, the cp genome has become prominent due to its conserved structure, maternal inheritance, and sufficient number of informative sites. In general, the cp genome of most land plants comes in a circular quadripartite structure consisting of three regions: one large single copy (LSC) and one small single copy (SSC), separated by two inverted repeats (IR) that mirror each other (Wicke et al., 2011). Although highly conserved, the cp genome sequences have highly variable sites at the interspecific level, which are potentially useful to be developed as DNA barcodes or specific markers that could distinguish closely related species apart (Daniell et al., 2021). As performing NGS is more cost-effective when compared to conventional Sanger sequencing, approximately 720 complete cp genome sequences of *Solanum* taxa have been published and deposited in the NCBI GenBank database to date (as of January 25, 2024). As nearly half of the published records come from redundant species, the current record is still far from the number of recognised species in *Solanum* and thus could not reveal a good representation of the genus. However, current findings have indicated that the complete cp genome sequence is promising in constructing a resolved phylogenetic tree within *Solanum* (Yan et al., 2022).

To provide more genetic information on some selected understudied *Solanum* species in China, in this study, we sequenced and characterised the complete cp genome sequence of *S. aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium*. Genome comparative and phylogenomic analyses are conducted using the complete cp genome sequences of

these three species of *Solanum* to reveal their phylogenetic relationship and evolutionary pattern within *Solanum*. The information presented in this study may contribute to the advancement of genetic and evolutionary studies for *Solanum* species.

Material and Methods

Plant materials, DNA extraction, and next-generation sequencing

Fresh leaves of *S. aculeatissimum*, *S. lasiocarpum*, *S. pittosporifolium*, were collected from natural populations in China (Table 1). The fresh leaf samples were kept in ziplock bags containing silica gels and transported back to the laboratory at Sun Yat-sen University for total genomic DNA extraction. Total genomic DNA extraction was carried out using the DNeasy Plant Mini Kit (Qiagen, USA) following the manufacturer's protocol. The DNA extract was quantified using a Qubit 4 fluorometer (Thermo Fisher Scientific, USA) for its purity and concentration prior to being sent to Guangzhou Jierui Biotechnology Company, Ltd. (Guangzhou, China) for next-generation sequencing. A 350-bp paired-end genomic library was prepared using the TrueSeq DNA Sample Prep Kit (Illumina, USA), and 150-bp paired-end reads were sequenced using the NovaSeq 6000 platform (Illumina, USA).

Table-1. Chloroplast genome features of the three species of *Solanum* used in this study.

Species	***S. aculeatissimum***	***S. lasiocarpum***	***S. pittosporifolium***
Place of origin	Qixi Mountain, Ji'an City, Jiangxi Province, China	Shenyi Village, Nanfeng Town, Danzhou City, Hainan Province, China	Dawu Mountain, Xinyi City, Guangdong Province, China
Collectors and collection number	Zhao W., Xu K., Liu Z., Liao F.; LXP-13-07963	Li R.; LRT-3207	Xu J., Chen J., Pan J., Liang W.; YKS-1782
GenBank accession number	PP234974	PP234975	PP234976
Genome length (bp)	155,821	155,671	155,926
GC content (%)	37.8	37.7	37.8
LSC (bp)	86,489	86,380	86,365
SSC (bp)	18,494	18,559	18,375
IR (bp)	25,419	25,366	25,593
Coding region (bp)	125,519	122,640	121,068
Non-coding region (bp)	30,302	33,031	34,858
Total number of genes	129	129	130
CDS	79	79	86
tRNA	37	37	37
rRNA	8	8	8

Chloroplast genome assembly, annotation, and visualisation

The raw NGS data was fed into the NOVOWrap v1.20 (Wu et al., 2021) pipeline to assemble the complete cp genome sequence by using the *rbcL* gene sequence of *Solanum virginianum* (GenBank accession no. MH046942) as the seed sequence. The total GC content of the assembled cp genome sequence was calculated using PhyloSuite (Zhang et al., 2020). Gene annotation and identification of the IR regions were carried out using GeSeq v2.03 (Tillich et al., 2017). The annotated cp genome sequence was manually checked for errors, and the physical map of the cp genome was visualised using OGDraw v1.3.1 (Greiner et al., 2019).

Repeat analyses

The MISA-web online software was used to calculate the occurrence of simple sequence repeats (SSRs) based on the parameter of minimum number of repeats set at 10, 6, 5, 5, 5, and 5 for mono-, di-, tri-, tetra-, penta-, and hexanucleotides (Beier et al., 2017). Large repeats, in the forms of palindromic, forward, reverse, and complement repeat sequences, were identified using the REPuter programme (Kurtz et al., 2001), with a hamming distance of 3 and a minimum repeat length of 30 bp.

Genome comparative and nucleotide divergence analyses

The IR border analysis was carried out for the three complete cp genome sequences assembled in this study. The genes adjacent to the IR borders were identified using CPJSdraw v1.0.0 (Li et al., 2023). By selecting the complete cp genome sequence of *S. nigrum* (GenBank accession no. MT621037) as the reference genome, genome comparative analysis was conducted using mVISTA (Frazer et al., 2004) for the three complete genome sequences used in this study. The cp genome sequences were aligned under Shuffle-LAGAN mode. Nucleotide diversity in the complete cp genomes of 24 selected species of *Solanum* was identified using DnaSP v5.0 (Librado and Rozas, 2009). The cp genome sequences are aligned using MAFFT v7.487 (Katoh and Standley, 2013) prior to analysis. A sliding window analysis was carried out based on a window length of 300 bp and a step size of 200 bp.

Phylogenetic analysis

Based on the work of Gagnon et al. (2021) and Yan et al. (2022), the phylogenetic tree of *Solanum* was reconstructed based on the complete cp genome sequences, with the sequence of IRA being excluded from analysis. Based on the availability of complete cp genome sequence data in the GenBank database, a total of 30 complete cp genome sequences from species of *Solanum* native to China, together with the three species used in this study, were included in the phylogenetic analysis. The cp genome sequences were aligned using MAFFT v7.487 (Katoh and Standley, 2013), while four closely related species, *Capsicum annuum* (GenBank accession no. JX270811), *C. baccatum* (GenBank accession no. OP374136), *Jaltomata bicolor* (GenBank accession no. MZ221901), and *J. sinuosa* (GenBank accession no. MZ221902), are included as outgroup species. Phylogenetic analysis was carried out using two methods, maximum likelihood (ML) and approximate Bayesian inference (aBI), using IQ-TREE v1.6.8 (Nguyen et al., 2015) embedded in the PhyloSuite programme (Zhang et al., 2020). Prior to phylogenetic tree reconstruction, the most optimum nucleotide substitution model for the cp genome dataset was calculated using ModelFinder (Kalyaanamoorthy et al., 2017). Based on the Bayesian inference criterion, the transversion model (TVM) under empirical based frequencies (+F), with invariable sites (+I) plus discrete Gamma model by default four rate categories (+G4) (=TVM+F+I+G4) was selected. The ML branch supports were obtained using the Shimodaira-Hasegawa approximate likelihood-ratio test (SH-aLRT) and the ultrafast bootstrapping algorithm (UFboot), based on 1,000 bootstrap replicates, while the aBI tree was calculated based on posterior probability (PP). The final tree was visualised using FigTree v1.4.4 (Rambaut et al., 2018).

Results

Chloroplast genome structure

Approximately 6 Gb of raw data was generated and used for cp genome assembly. All three complete cp genome sequences showed a quadripartite structure that consists of an LSC, an SSC, and a pair of IR regions (Figure 1). The complete cp genome sizes of *S. aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium* were 155,821 bp, 155,671 bp, and 155,926 bp, respectively. Among the three species, *S. aculeatissumum* has the largest LSC region (86,489 bp); *S. lasiocarpum* has the largest SSC region (18,559 bp) but the smallest IR region (25,336 bp); *S. pittosporifolium* has the smallest LSC (86,365 bp) and SSC regions (18,375 bp) but the largest IR region (25,593 bp). A total of 129 genes were annotated in complete cp genomes of *S. aculeatissimum* and *S. lasiocarpum*, including 84 CDS, 37 tRNA, and eight rRNA genes (Table 2). *S. pittosporifolium* has 130 total genes annotated, which includes an additional *clpP1* gene. Among these genes, 18 were duplicated in the IR region, including *ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *rrn4.5*, *rrn5*, *rrn16*, *rrn23*, *trnA*-UGC, *trnI*-CAU, *trnI*-GAU, *trnL*-CAA, *trnN*-GUU, *trnR*-ACG, *trnV*-GAC, *ycf*1, and *ycf*2. For the CDSs, seven of them contained one intron (i.e., *atp*F, *clpP1*, *ndh*A, *ndhB*, *petB*, *petD*, *rpl2*, and *rpoC1*), while one contained two introns (i.e., *paf1*). The total GC content of all the complete cp

genome sequences was 37.8%, except for *S. lasiocarpum*, which was 37.7%. All three assembled genomes were submitted to GenBank under accession numbers PP234974, PP234975, and PP234976 for *S. aculeatissimum* and *S. lasiocarpum* and *S. pittosporifolium*, respectively.

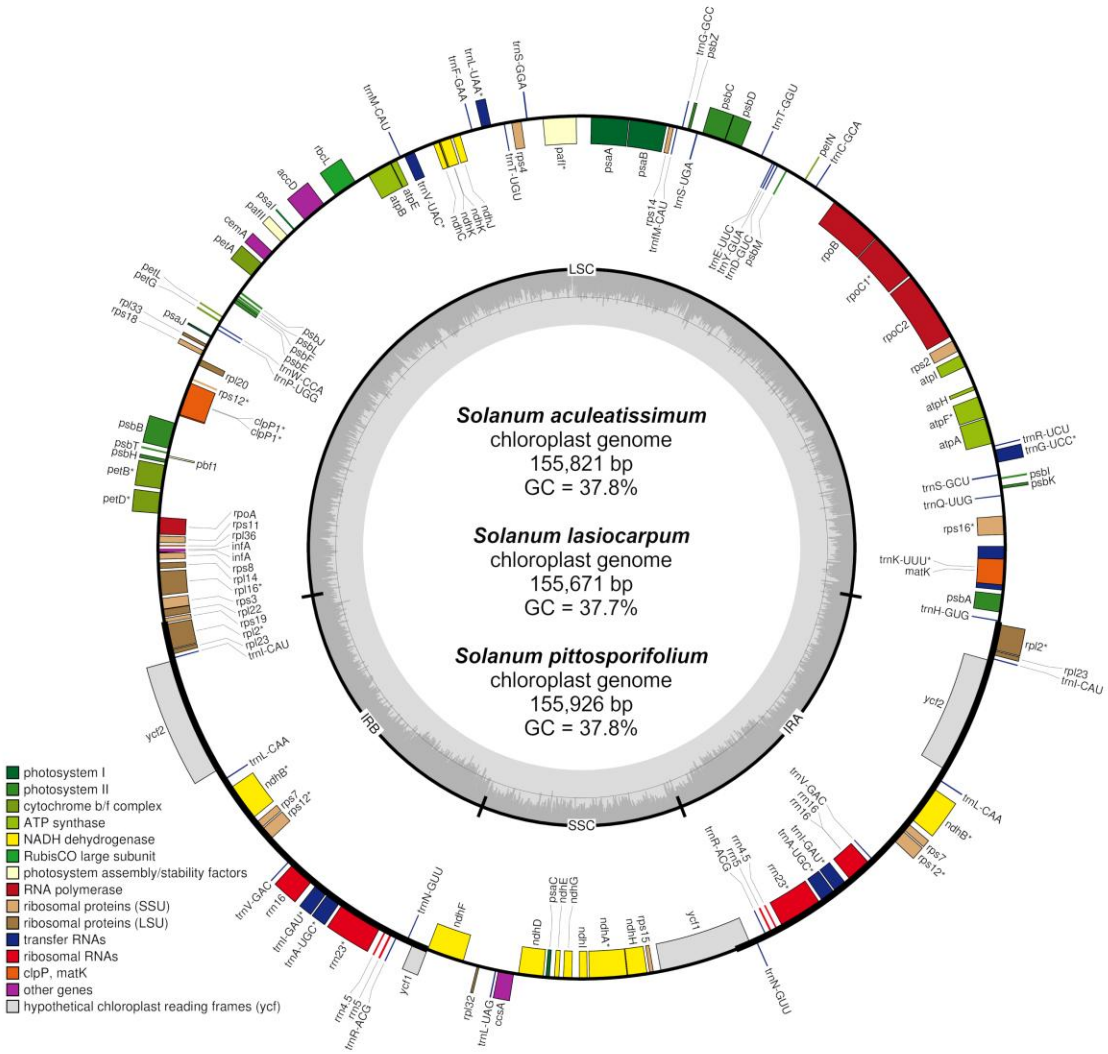

Figure-1. Chloroplast genome map of three species of *Solanum*. Genes annotated in the circle are transcribed counterclockwise, while the genes annotated outside the circle are transcribed clockwise. The information of each species including genome length and GC content are indicated below the species name, at the center of the genome map.

Table-2. List of annotated genes in the chloroplast genome of *S. aculeatissimum*, *S. lasiocarpum* and *S. pittosporifolium*, along with their group and function. Genes that are duplicated in the inverted repeat region are indicated with an asterisk (*).

Function	Gene Group	Name of gene
Photosynthesis pathway	Photosystem I	*psaA*, *psaB, psaC*, *psaI*, *psaJ*
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbT*, *psbZ*
	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*
	NADH complex	*ndhA*, *ndhB**, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
	Cytochrome b6/f complex	*petA*, *petB*, *petD*, *petG*, *petL*, *petN*
	Large subunit of Rubisco	*rbcL*
	Photosystem biogenensis	*pbf1*
Structural RNAs	Transfer RNAs	*trnA*-UGC*, *trnC*-GCA, *trnD*-GUC, *trnE*-UUC, *trnF*-GAA, *trnfM*-CAU, *trnG*-GCC, *trnG*-UCC, *trnH*-GUG, *trnI*-CAU*, *trnI*-GAU*, *trnK*-UUU, *trnL*-CAA*, *trnL*-UAA, *trnL*-UAG, *trnM*-CAU, *trnN*-GUU*, *trnP*-UGG, *trnQ*-UUG, *trnR*-ACG*, *trnR*-UCU, *trnS*-GCU, *trnS*-GGA, *trnS*-UGA, *trnT*-GGU, *trnT*-UGU, *trnV*-GAC*, *trnV*-UAC, *trnW*-CCA, *trnY*-GUA
	Ribosomal RNAs	*rrn4.5**, *rrn5**, *rrn16**, *rrn23**
Genetic apparatus	Large subunit of ribosomal protein	*rpl2**, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23**, *rpl32*, *rpl33*, *rpl36*
	Small subunit of ribosomal protein	*rps2*, *rps3*, *rps4*, *rps7**, *rps8*, *rps11*, *rps12**, *rps14*, *rps15*, *rps16*, *rps18*, *rps19*
	Subunits of RNA polymerase	*pafI*, *pafII*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*
Others	Maturase	*matK*
	Protease	*clpP1*[#]
	Inner envelope membrane	*cemA*
	Cytochrome biogenesis protein	*ccsA*
	Fatty Acid synthesis	*accD*
Unknown	Open-reading frame	*ycf1**, *ycf2**

only present in *Solanum pittosporifolium*.

Simple sequence repeats and large repeats

The complete cp genome sequence of *S. aculeatissimum* had the most SSRs identified (n = 42), while *S. pittosporifolium* had the least (n = 33) (Figure 2a). The mononucleotide repeat A/T was the dominant repeat unit in all three cp genome sequences. Trinucleotide repeats were identified in *S. aculeatissimum* and *S. pittosporifolium*, but not in *S. lasiocarpum*. The mononucleotide repeat unit C/G and trinucleotide repeat unit AAC/GTT were unique to *S.*

aculeatissimum, which both were accounted for with one count.

A total of 37, 43, and 30 large repeats were identified in the complete cp genomes of *S. aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium*, respectively (Figure 2b). All three species had large repeats in the form of palindromic and forward, while the presence of reverse and complementary large repeats was only recorded in *S. lasiocarpum*. When compared to palindromic large repeats, *S. aculeatissimum* and *S. lasiocarpum* accounted for a higher count for their forward large repeats, while S. *pittosporifolium* had an equal count of both palindromic and forward large repeats.

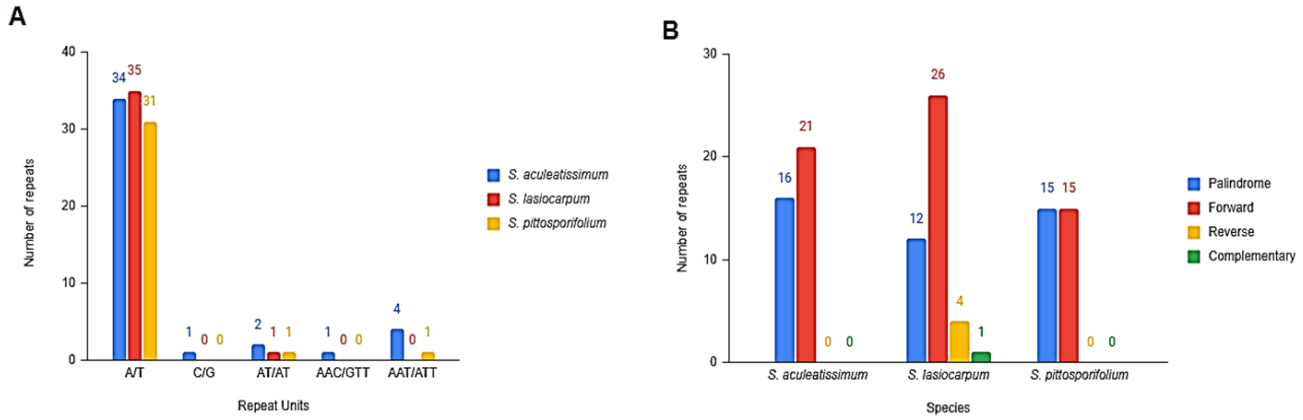

Figure-2. Repeat analyses of the complete chloroplast genome sequence of *S. aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium*. (a) Simple sequence repeats based on the types of repeats, (b) types of large repeats, including palindromic, forward, reverse, and complementary.

Comparison of IR borders

The three species of *Solanum* used in this study showed the same gene content that was adjacent to the IR borders (Figure 3). At the junction between the LSC and IRB regions (JLB), *rpl2* was placed in the IRB region, while *rps19* was crossing over the border into the LSC region. For the junction between the SSC and IRB regions (JSB), the *ndh*F gene was found in the SSC region, while the *ycf1* gene was extending from the IRB region into the SSC region. A similar incident occurred for the other copy of *ycf1*, in which it was found crossing over the junction between the SSC and IRA regions (JSA) into the SSC region. The *trnN* gene was at least 1,200-bp away from JSA, while the *rpl*2 gene was at least 140-bp away from the junction between the LSC and IRA regions (JLA), both placed in the IRA region. The *trnH* gene was in the LSC region, at least 75-bp away from JLA.

Comparison of genome nucleotide variation

By comparing the reference genome of *S. nigrum*, a total of five, six, and two distinct gaps were detected in the complete cp genome sequences of *S. aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium*, respectively (Figure 4). Based on the genome alignment, the first distinct gap was found at the non-coding region of *rps16-trnQ*-UUG in the cp genome of *S. lasiocarpum*. The second distinct gap was found in the cp genome of *S. pittosporifolium*, in the non-coding region between *psbI* and *psbK*. The cp genomes of *S. aculeatissimum* and *S. lasiocarpum* had five distinct gaps that were located in similar regions: in the non-coding regions between *psbM* and *trnD*-GUC, *trnE*-UUC and *trnT*-GGU, as well as *trnL*-UAA and *trnF*-GAA, and in the coding regions of both the *ycf1* genes. A distinct gap was also found in the intron region of the *ndhA* gene of *S. pittosporifolium*.

https://doi.org/10.35495/ajab.2025.042

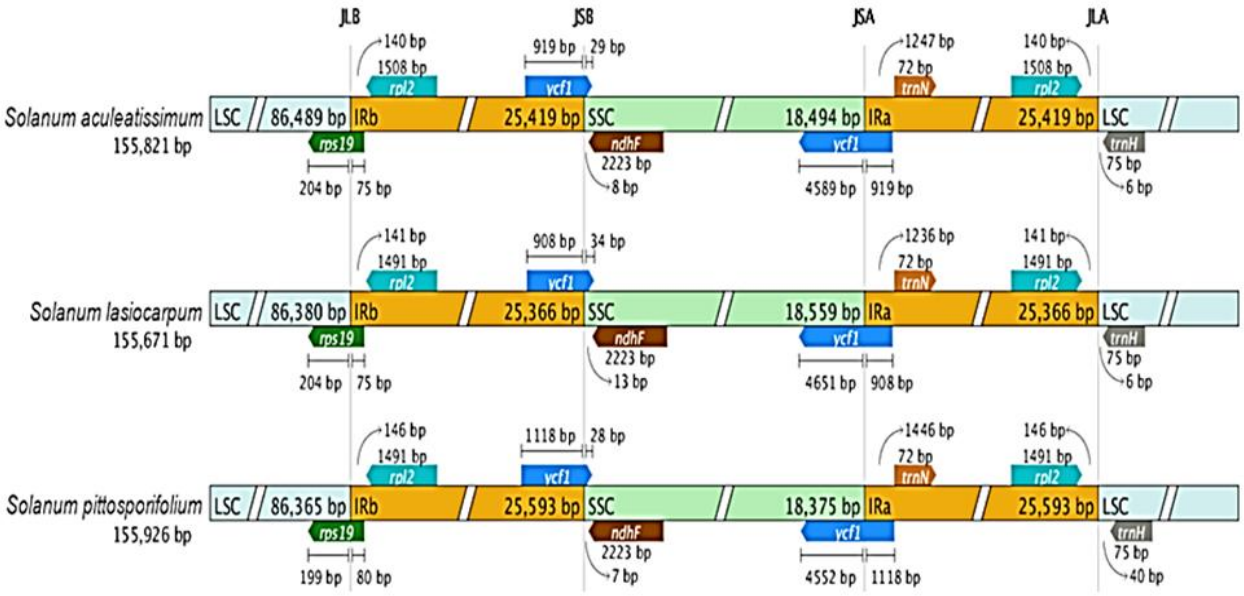

Figure-3. Inverted repeats border comparison between the complete chloroplast genome of three species of *Solanum*.

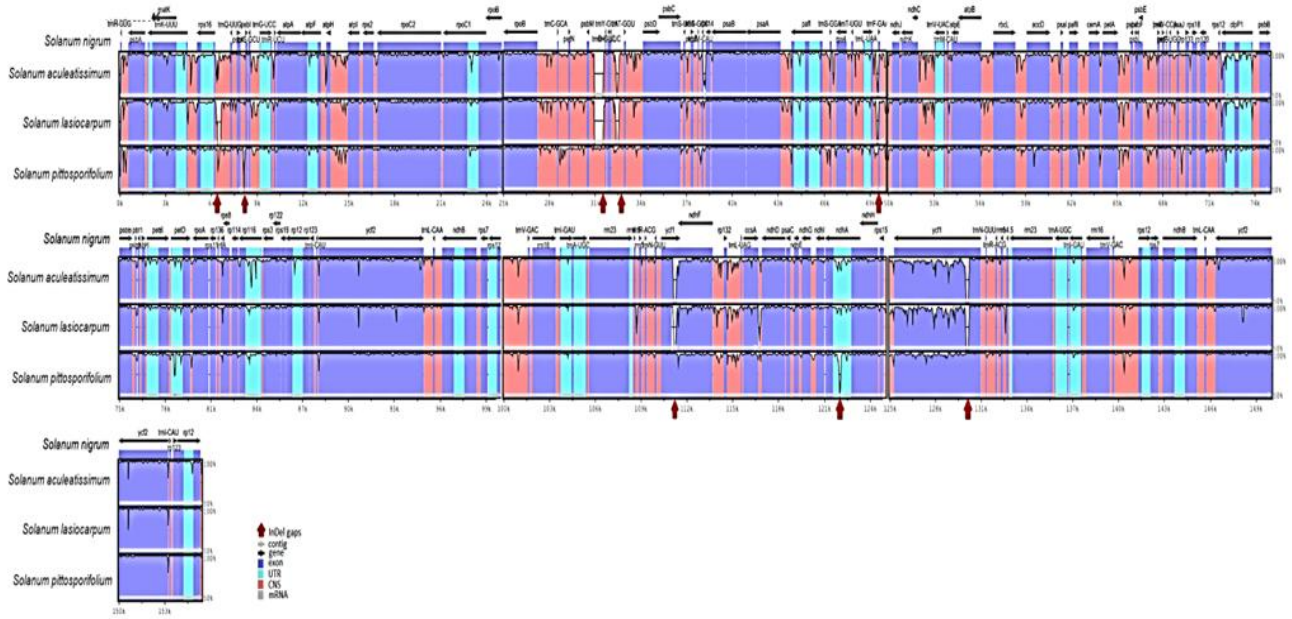

Figure-4. Comparison of chloroplast genome nucleotide variation of three species of *Solanum* using mVISTA. The complete chloroplast genome sequence of *S. nigrum* (GenBank accession no. MT621037) is used as the reference genome. The y-axis represents the percent identity within 50–100%. Grey arrows indicate the direction of gene transcription. Red arrows indicate the distinct gaps detected in the genome alignment.

https://doi.org/10.35495/ajab.2025.042

When Pi > 0.03, at least four highly variable regions were detected in the genome alignment of the 33 species of *Solanum* (Figure 5). These highly variable regions include the intergenic spacer regions *ndhC-trnV*-UAC and *rpl13-rps18*, both of which are in the LSC region, as well as the coding regions of *petD* and *ycf1*, which are in the LSC and IRA regions, respectively.

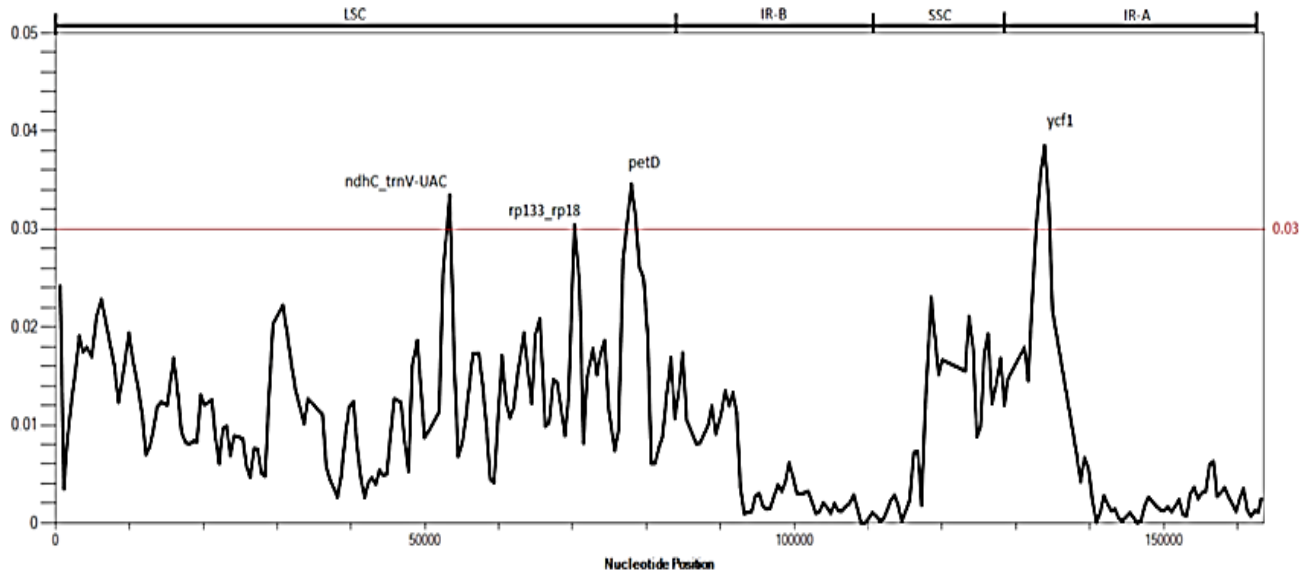

Figure-5. Nucleotide variability (Pi) analysis based on the complete chloroplast genome sequence alignment of 24 species of *Solanum*, including the three used in this study. The list of the 21 species of *Solanum* used in this analysis is presented in Appendix A.

Phylogenetic inference

Based on the current sampling size, the ML tree showed a well-resolved relationship within *Solanum*; a branch node that came with a support value based on the SH-aLRT (left), PP (middle), and UFboot (right) are considered reliable when the support value is ≥ 80%, ≥ 0.95, and ≥ 95%, respectively (Figure 6). From the ML tree, *S. pittosporifolium* is clustered with *S. dulcamara* and *S. septemlobum*, in which *S. dulcamara* was the first to diverge. *Solanum aculeatissimum* and *S. lasiocarpum* were clustered with *S. capsicoides*, and the three of them form a clade that is sister to another clade containing 11 other species of *Solanum*.

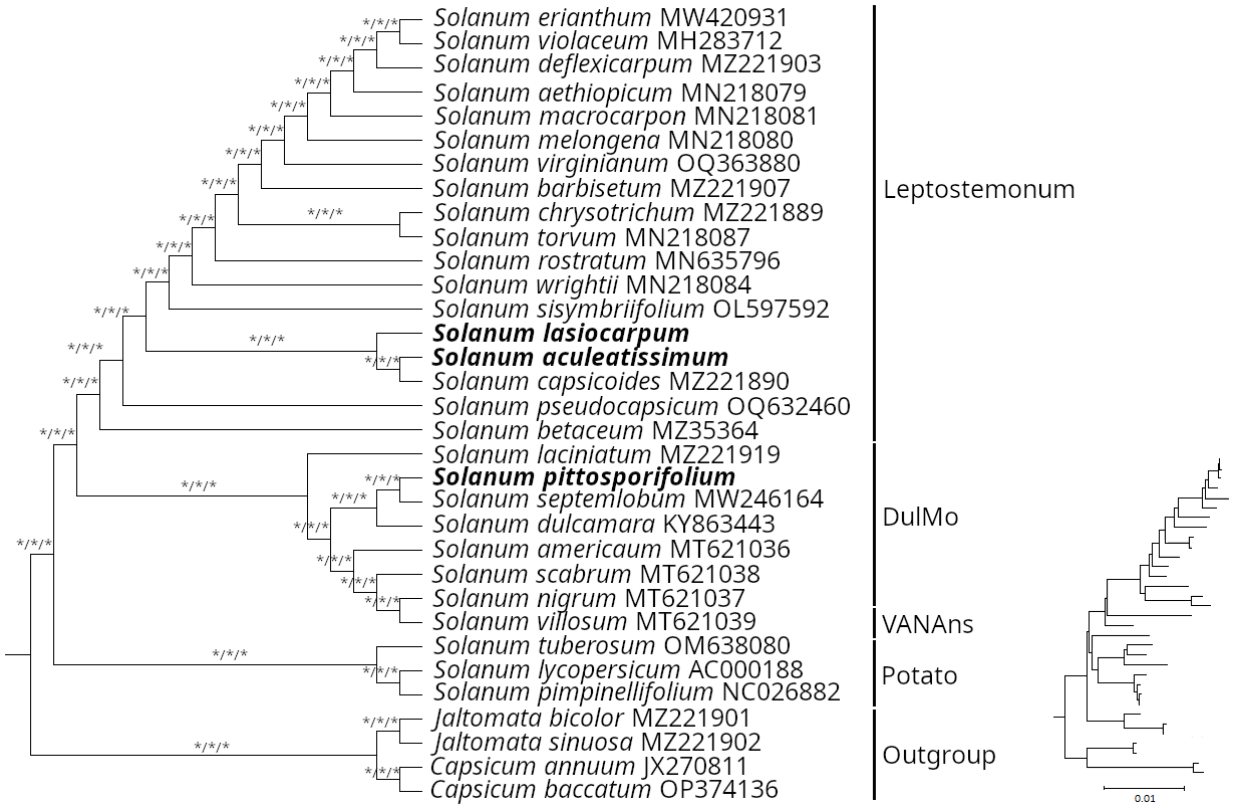

Figure-6. Phylogenetic tree based on the complete chloroplast genome sequence of 29 *Solanum* species. The sequence of IRA was removed and four closely related species, *Jaltomata sinuosa* (GenBank accession no. MTMZ221902), *Jaltomata bicolor* (GenBank accession no. MZ221901), *Capsicum annuum* (GenBank accession no. JX270811), and *Capsicum baccatum* (GenBank accession no. OP374136) were included as outgroup. Bootstrap support (BS; left), posterior probability (PP; middle), and ultrafast bootstrap (UFboot; right) that are considered reliable (BS ≥75%, PP ≥0.95, UFboot ≥ 95%) are indicated with an asterisk (*) above the branch.

Discussion

The complete cp genomes of *Solanum aculeatissimum*, *S. lasiocarpum*, and *S. pittosporifolium* displayed a structure that is similar to that of most angiosperms, which is a quadripartite circular structure (Xu et al., 2025). The cp genome size of the three species of *Solanum* used in this study is similar to other published cp genomes of *Solanum*, of which they are greater than that of *S. elaeagnifolium* (155,049 bp; Zhu et al., 2020) and *S. rostratum* (155,296 bp; Shi and Qiu, 2020) and smaller than that of *S. betacea* (155,937 bp; Li et al., 2021) and *S. macrocarpon* (155,937 bp; Yang et al., 2023). The three cp genomes annotated in this study had the similar number of total genes annotated, i.e., 129 to 130. When compared to other published complete cp genomes of *Solanum*, the total number of genes would fall between 129 and 134 (Li et al., 2021; Yang et al., 2023). The difference in the number of annotated genes could be a variation in evolutionary pattern among these species; however, we do not exclude the possibility that the use of different annotation tools would also affect the effective-ness of the annotation, which would result in a different number of annotated gene (Guyeux et al., 2019). The total GC content of the three species of *Solanum* was somewhat congruent to those of published cp genomes of *Solanum*, which is generally between 37.7 % and 37.9 % (Yang et al., 2023). The low GC content in the cp genome of *Solanum* is contributed by the high count of A/T nucleotides in the genome sequence. In general, the cp genome sequences of most land plants are A/T-biassed, while most of them occur in the non-coding regions (Kang et al., 2023).

Sequence repeats such as SSR and large repeats are generally used as molecular markers linked to desirable phenotypic traits; they also play an important role in evolution by rearranging, recombining, and inverting genome structure (Kumar, 1999). In this study, we found that more than 80% of the SSRs identified in the three cp genomes were A/T-derived mononucleotide repeats. A similar finding was reported for a study on 29 tomato (*S. lycopersicum*) germplasms, in which more than 97% of SSRs were mononucleotide repeats, while the di- and trinucleotides consisted of the A/T motif (Wang et al., 2023). The presence of palindromic and forward large repeats is common in the cp genomes of *Solanum*, while reverse and complementary large repeats are rare in general (Amiryousefi et al., 2018; Yang et al., 2023). In a recent study by Yang et al. (2023), reverse large repeats were found in at least five species of *Solanum*, including *S. aethiopium*, *S. macrocarpon*, *S. wrightii*, *S. indicum*, and *S. sisymbriifolium*, but only *S. aethiopium* was recorded to contain complementary large repeats. Based on this limited information, we could only speculate that the occurrence of reverse and complimentary large repeats in the cp genomes of *Solanum* could have happened independently among its species.

Genome comparative analyses revealed that the gene arrangement and IR borders were conserved among the three species of *Solanum*. Although small variations, such as the distance between the IR junctions and the genes adjacent to them, are noticeable between the three species of *Solanum*, the placement of the IR junction is somewhat consistent in the cp genomes. The differences in the number of sites between the IR junction and the adjacent gene are likely due to the expansion and contraction of the IR regions. The expansion and contraction of inverted regions are known to be the main cause of gaining or losing genes and are considered a common phenomenon for evolution (Zhu et al., 2016). When compared to the cp genome sequence of *S. nigrum*, the nucleotide variation in the cp genomes of the three species of *Solanum* displayed slight variations. The distinct gaps, when further confirmed through the genome sequence alignment, are the presence of insertions and deletions (indels) of bases. Among the three cp genome sequences, the longest indel was found in the LSC region of *S. aculeatissimum* and *S. lasiocarpum*, between the genes *psbM* and *trnD*-GUC, which was at least 580 bp long. This was followed by the second longest indel in the cp genome of *S. aculeatissimum* and *S. lasiocarpum*, which was at least 200 bp long and was in the IRA region. Indels are generally found in the LSC region in cp genomes, and the length of the insertions and deletions are usually the reason for changes in genome length (Niu et al., 2017), while they can also be further developed into specific markers useful in genotyping and species identification, as demonstrated by Lee et al. (2021). Based on the nucleotide diversity analysis, we revealed that highly variable regions were mostly concentrated in the LSC region, of which other studies also mentioned the same (Niu et al., 2017). This is because the LSC region is the largest in size, making up of more than half of the cp genome when compared to the other regions in the cp genome. One of these highly variable regions, the *ycf*1 gene, was once proposed as a useful gene region that could be used as a DNA barcode (Dong et al., 2015). Despite the fact that the NGS service has become affordable in many countries, the effort to sequence the cp genomes of *Solanum* would be costly as *Solanum* is a diverse genus. Therefore, it would be a cost-effective approach to design useful DNA primers that could amplify and provide sufficient information to delimit all species through conventional Sanger sequencing techniques. To develop such primers, the highly variable regions identified in the cp genome sequence could be further analysed and tested for their efficacy in delimiting closely related species apart, as demonstrated by Hishamuddin et al. (2023).

The classification of *Solanum* has been a challenging task for taxonomists over the years (Fawzi and Habeeb, 2016), while with the aid of molecular tools, researchers have proposed new classification methods to comb out the complex relationships in the genus. The molecular classification of *Solanum* started by using a single chloroplast gene sequence, *ndhF* (Bohs, 2005), to the use of multi-loci datasets that gradually elevated the resolution of the phylogenetic relationship within *Solanum* (Weese and Bohs, 2007; Stern et al., 2011; Stern and Bohs, 2012; Tepe et al., 2016; Särkinen et al., 2008). The first large scale study on the phylogenetic analysis of *Solanum* using the cp genome dataset revealed a cyto-nuclear discordance in *Solanum* (Gagnon et al., 2021). Still, the taxonomic framework used in *Solanum* dividing the large genus into major and minor clades based on genome-scale datasets is considered robust; at present, it is generally accepted that *Solanum* is divided into 12 major clades, i.e., Brevantherum, Clandestinum-Mapiriense, Cyphomandra, DulMo, Geminata, Leptostemonum,

Nemorense, Potato, Regmandra, Thelopodium, VANAns, and Wendlandii-Allophyllum (Särkinen et al., 2008; Gagnon et al., 2021). Based on this classification method, the cp genome-based ML tree showed that both the species *S. aculeatissimum* and *S. lasiocarpum* are grouped under the Leptostemonum clade, while *S. pittosporifolium* is grouped under the DulMo clade. Species that are under the Leptostemonum clade are characterised by the presence of the spiny stem and leaf, and the members of the DulMo clade are usually in the form of vines (Knapp, 2013; Aubriot and Knapp, 2022). The molecular placement of the three species of *Solanum* used in this study based on the complete cp genome sequences is also congruent with the findings by Särkinen et al. (2008) that used the combined dataset of ITS, *matK*, *ndhF*, *psbA-trnH*, *trnL-trnF*, *trnS-trnG*, and *waxy*.

Conclusions

In this study, the complete cp genomes of three species of *Solanum* were sequenced, assembled, and annotated, contributing new genomic data for understudied species within the genus. Genome comparative analyses revealed that the cp genome structure and gene content of all three species were highly conserved, consistent with previously reported patterns in *Solanum*. Phylogenetic analysis based on the cp genome sequences provided initial insights into species relationships; however, the limited sampling restricts the broader applicability of the findings. Therefore, the inferred relationships should be considered provisional. Despite these limitations, the genomic information presented in this study serves as a useful resource for future studies on the taxonomy, phylogeny, and evolutionary dynamics of *Solanum*.

Acknowledgements

The first author is a recipient of the INTI International University Graduate Research Assistant grant scheme.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: This research was funded by INTI International University Seed Grant Scheme, grant number INTI-FHLS-11-03-2022; Foundation of Qingyuan Forestry Bureau, grant number HT-99982024; Guangdong Provincial Ecological Forestry Development Project, grant number 2020141; Hainan Province Major Science and Technology Plan Project, grant number ZDZX2021034.

Contribution of Authors

Ng RJY: Formal analysis, software and original draft preparation.

Chen Z: Sample collection, formal analysis, software and original draft preparation.

Tan L: Formal analysis, software and original draft preparation.

Law D: Project administration, supervision, manuscript review and editing.

Liao: Sample collection, resources, manuscript review and editing.

Liu Y: Conceptualization, validation, manuscript review and editing.

Lee SY: Conceptualization, resources, supervision, manuscript review and editing.

References

Amiryousefi A, Hyvönen J and Poczai P, 2018. The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): Plastid genome structure evolution in Solanaceae. PLoS One 13:e0196069. DOI: 10.1371/journal.pone.0196069.

Aubriot X and Knapp S, 2022. A revision of the "spiny solanums" of Tropical Asia (*Solanum*, the Leptostemonum Clade, Solanaceae). PhytoKeys 198:1-270. DOI: 10.3897/phytokeys.198.79514.

Beier S, Thiel T, Münch T, Scholz U and Mascher M, 2017. MISA-web: A web server for microsatellite prediction. Bioinform. 33:2583-2585, DOI: 10.1093/bioinformatics/btx198.

Bohs L, 2005. Major clades in *Solanum* based on *ndhF* sequence data, pp. 24-49. In A festschrift for William G. D'Arcy: The legacy of a taxonomist. Missouri Botanical Garden Press, Mo.

Daniell H, Jin S, Zhu X, Gitzendanner MA, Soltis DE and Soltis PS, 2021. Green giant – a tiny chloroplast genome with mighty power to produce high-value proteins: history and phylogeny. Plant Biotechnol. J. 19:430-447. DOI: 10.1111/pbi.13556.

D'Arcy WG, 1972. Solanaceae studies II: Typification of subdivisions of *Solanum*. Ann. Missouri Bot. Gard. 59:262-278. DOI: 10.2307/2394758.

Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, Cheng T, Guo J and Zhou S, 2015. Ycf1, the most

promising plastid DNA barcode of land plants. Sci. Rep. 5:8348. DOI: 10.1038/srep08348.

Fawzi NM and Habeeb HR, 2016. Taxonomic study on the wild species of genus *Solanum* L. in Egypt. Ann. Agric. Sci. 61:165-173. DOI: 10.1016/j.anas.2016.10.003.

Frazer KA, Patcher L, Poliakov A, Rubin EM and Dubchak I, 2004. VISTA: Computational tools for comparative genomics. Nucleic Acids Res. 32(suppl_2):W273-W279. DOI: 10.1093/nar/gkh458.

Gagnon E, Hilgenhof R, Orejuela A, McDonnel A, Sablok G, Aubriot X, Giacomin L, Goubea Y, Bragionis T, Stehmann JR, Bohs L, Dodsworth S, Martine C, Poczai P, Knapp S and Särkinen T, 2021. Phylogenomic discordance suggests polytomies along the backbone of the large genus *Solanum*. Am. J. Bot. 109:580-601. DOI: 10.1002/ajb2.1827.

Gebhardt C, 2016. The historical role of species from the Solanaceae plant family in genetic research. Theor. Appl. Genet. 129:2281-2294. DOI: 10.1007/s00122-016-2804-1.

Greiner S, Lehwark P and Bock R, 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47(W1): W59-W64. DOI: 10.1093/nar/gkz238.

Guyeux C, Charr JC, Tran HT, Furtado A, Henry RJ, Crouzillat D, Guyot R and Hamon P, 2019. Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. PLoS One 14:e0216347. DOI: 10.1371/journal.pone.0216347.

Hishamuddin MS, Lee SY, Syazwan SA, Ramlee SI, Lamasudin DU and Mohamed R, 2023. Highly divergent regions in the complete plastome sequences of *Aquilaria* are suitable for DNA barcoding applications including identifying species origin of agarwood products. 3 Biotech 13:78. DOI: 10.1007/s13205-023-03479-1.

Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A and Jermiin LS, 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14:587-589. DOI: 10.1038/nmeth.4285.

Kang J, Giang VNL, Park H, Park YS, Cho W, Nguyen VB, Shim H, Waminal NE, Park JY, Kim HH and Yang T, 2023. Evolution of the Araliaceae family involved rapid diversification of the Asian Palmate group and Hydrocotyle specific mutational pressure. Sci. Rep. 13:22325. DOI: 10.1038/s41598-023-49830-7.

Katoh K and Standley DM, 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772-780. DOI: 10.1093/molbev/mst010.

Kaunda JS and Zhang YJ, 2019. The genus *Solanum*: An ethnopharmacological, phytochemical and biological properties review. Nat. Prod. Bioprospect. 9:77-137. DOI: 10.1007/s13659-019-0201-6.

Knapp S, 2013. A revision of the Dulcamaroid clade of *Solanum* L. (Solanaceae). Phytokeys 22:1-432, DOI:10.3897/phytokeys.22.4041.

Kumar LS, 1999. DNA markers in plant improvement: an overview. Biotechnol. Adv. 17:143-182. DOI: 10.1016/s0734-9750(98)00018-4.

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J and Giegerich R, 2001. REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 29:4633-4642. DOI: 10.1093/nar/29.22.4633.

Lee SY, Chen Z, Chen J, Fan Q, Liu J and Liao W, 2021. Authentication of medicinal herb *Wikstroemia indica* using novel DNA markers derived from the chloroplast genome sequences. J. Appl. Res. Med. Aromat. Plants 24:100315, DOI: 10.1016/j.jarmap.2021.100315.

Levin RA, Myers NR and Bohs L, 2006. Phylogenetic relationships among the "spiny solanums" (*Solanum* subgenus Leptostemonum, Solanaceae). Am. J. Bot. 93:157-169. DOI: 10.3732/ajb.93.1.157.

Li H, Guo Q, Xu L, Gao H, Liu L and Zhou X, 2023. CPJSdraw: Analysis and visualization of junction sites of chloroplast genomes. PeerJ 11:e15326. DOI: 10.7717/peerj.15326

Li S, Zheng X, Duan H and Dong Q, 2021. The complete chloroplast genome of *Solanum betacea* (Solanaceae, Solaneae). Mitochondrial DNA B 6:1642-1644. DOI: 10.1080/23802359.2021.1927219.

Librado P and Rozas J, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinform. 25:1451-1452. DOI: 10.1093/bioinformatics/btp187.

Musarella CM, 2020. *Solanum torvum* Sw. (Solanaceae): A new alien species for Europe.

Genet. Resour. Crop Evol. 67:515-522. DOI:10.1007/s10722-019-00822-5.

Nguyen LT, Schmidt HA, Von Haeseler A and Minh BQ, 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32:268-274. DOI: 10.1093/molbev/msu300.

Niu Z, Zhu S, Pan J, Li L, Sun J and Ding X, 2017. Comparative analysis of *Dendrobium* plastomes and utility of plastomic mutational hotspots. Sci. Rep. 7:2073. DOI: 10.1038/s41598-017-02252-8.

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF and Collier SM, 2008. A molecular phylogeny of the Solanaceae. Taxon 57:1159-1181. DOI: 10.1002/tax.574010.

POWO, 2024. Plants of the World Online. http://www.plantsoftheworldonline.org/ (accessed 25 January 2024.)

Rambaut A, Drummond AJ, Xin D, Baele G and Suchard MA, 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67:901-904. DOI: 10.1093/sysbio/syy032.

Särkinen T, Poczai P, Barboza GE, van der Weerden, GM, Baden M and Knapp SA, 2008. Revision of the Old World black night-shades (Morelloid clade of *Solanum* L., Solanaceae). PhytoKeys 106:1-223. DOI: 10.3897/phytokeys.106.21991.

Shi X and Qiu J, 2020. The complete chloroplast genome sequence of an invasive plant *Solanum rostratum* (Solanaceae). Mitochondrial DNA B 5:626-628. DOI: 10.1080/23802359.2020.1714506.

Stern S and Bohs L, 2012. An explosive innovation: Phylogenetic relationships of *Solanum* section Gonatotrichum (Solanaceae). PhytoKeys 8:89-98. DOI: 10.3897/phytokeys.8.2199.

Stern S, Agra MF and Bohs L, 2011. Molecular delimitation of clades within New World species of the "spiny solanums" (*Solanum* subg. Leptostemonum). Taxon 60:1429-1441. DOI: 10.1002/tax.605018.

Tepe EJ, Anderson GJ, Spooner DM and Bohs L, 2016. Relationships among wild relatives of the tomato, potato, and pepino. Taxon 65:262-276. DOI: 10.12705/652.4.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock and Greiner S, 2017, GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1):W6-W11. DOI: 10.1093/nar/gkx391.

Tynkevich YO, Shelyfist AY, Kozub LV, Hemleben V, Panchuk II and Volkov RA, 2022. 5S ribosomal DNA of genus *Solanum*: Molecular organisation, evolution, and taxonomy. Front. Plant Sci. 13:852406. DOI: 10.3389/fpls.2022.852406.

Wang X, Bai S, Zhang Z, Zheng F, Song L, Wen L, Guo M, Cheng G, Yao W, Gao Y and Li J, 2023. Comparative analysis of chloroplast genomes of 29 tomato germplasms: genome structures, phylogenetic relationships, and adaptive evolution. Front. Plant Sci. 14:1179009. DOI:10.3389/fpls.2023.1179009.

Weese TL and Bohs LA, 2007. Three-gene phylogeny of the genus *Solanum* (Solanaceae). Syst. Bot. 32:445-463. DOI:10.1600/036364407781179671.

Wicke S, Schneeweiss GM, Depamphilis CW, Müller KF and Quandt D, 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol. Biol. 76:273-297. DOI: 10.1007/s11103-011-9762-4.

Wu P, Xu C, Chen H, Yang J, Zhang X and Zhou S, 2021. NOVOWrap: an automated solution for plastid genome assembly and structure standardization. Mol. Ecol. Resour. 21:2177-2186. DOI: 10.1111/1755-0998.13410.

Xu KW, Yang Y, Chen H, Lin CX, Jiang L, Guo ZL, Li M, Hao MZ and Meng KK, 2025. Extensive cytonuclear discordance revealed by phylogenomic analyses suggests complex evolutionary history in the holly genus *Ilex* (Aquifoliaceae). Mol. Phylogenet. Evol. 204:108255. DOI: 10.1016/j.ympev.2024.108255.

Yan L, Zhu Z, Wang P, Fu C, Guan X, Kear P, Zhang C and Zhu G, 2022. Comparative analysis of 343 plastid genomes of *Solanum* section Petota: Insights into potato diversity, phylogeny, and species discrimination. J. Syst. Evol. 61:599-612. DOI: 10.1111/jse.12898.

Yang Q, Li Y, Cai L, Gan G, Wang P, Li W, Li W, Jiang Y, Li D, Wang M, Xiong C, Chen R and Wang Y, 2023. Characteristics, comparative analysis, and phylogenetic relationships of chloroplast genomes of cultivars and wild relatives of eggplant (*Solanum melongena*).

Curr. Issues Mol. Biol. 45:2832-2846. DOI: 10.3390/cimb45040185.

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX and Wang GT, 2020. PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol. Ecol. Resour. 20:348-355. DOI: 10.1111/1755-0998.13096.

Zhu A, Guo W, Gupta S, Fan W and Mower J, 2016. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol. 209:1747-1756. DOI:10.1111/nph.13743.

Zhu X, Wang A, Wu H and Lin J, 2020. Chloroplast genome of silverleaf nightshade (*Solanum elaeagnifolium* Cav.), a weed of national significance in Australia. Mitochondrial DNA B 5:2477-2479. DOI: 10.1080/23802359.2020.1775527.