Complete chloroplast genome sequence and structural analysis of the medicinal plant *Lycium chinense* Mill

Zerui Yang [1], **Yuying Huang** [1], **Xiasheng Zheng** [1], **Song Huang** Corresp., [1], **Lingling Liang** Corresp. [2]

[1] Mathematical Engineering Academy of Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China

[2] Pharmaceugtical school, YouJiang Medical University For Nationalities, Baise, Guangxi, China

Corresponding Authors: Song Huang, Lingling Liang
Email address: huangnn421@163.com, linglingyyzz@126.com

Lycium chinense Mill, an important Chinese herbal medicine, is emphasized as a healthy food and is widely used as a dietary supplement. Here we sequenced and analyzed the complete chloroplast (CP) genome of the *L. chinense*, which is 155,756 bp in length and with 37.8% GC content. This CP genome consists of a pair of inverted repeat regions (IRa and IRb) of 25,476 bp, separated by a large single-copy region (LSC) and a small single-copy region (SSC), with length of 86,595 and 18,209 bp, respectively. Annotation results revealed that the *L. chinense* CP genome contains 114 genes, 16 of which are duplicated genes. Most of the 85 protein-coding genes have a usual ATG start codon, except for 3 genes including *rps12*, *psbL* and *ndhD*. Furthermore, most of the simple sequence repeats (SSRs) are short polyadenine or polythymine repeats that contribute to the high AT content of the chloroplast genome. Revealing of the complete sequences and annotation of the *L. chinense* chloroplast genome will facilitate phylogenic, population and genetic engineering research investigations involving this particular species.

Complete chloroplast genome sequence and structural analysis of the medicinal plant *Lycium chinense* Mill

Zerui Yang[1], Yuying Huang[1], Xiasheng Zheng[1], Song Huang[1], Lingling Liang[2]

[1]Mathematical Engineering Academy of Chinese Medicine, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China
[2] Pharmaceugtical school, YouJiang Medical University For Nationalities, Baise, Guangxi, China

Corresponding Author:
Lingling Liang[2]
98th Chengxiang Road, Youjiang District, Baise, Guangxi, 533000, China
Email address: linglingyyzz@126.com
Song Huang[1]
232th Waihuangdong Road, Higher Education Mega Center, Panyu District
Email address: huangnn421@163.com

Abstract

Lycium chinense Mill, an important Chinese herbal medicine, is widely used as a dietary supplement and functional food. Here the chloroplast (CP) genome of *L. chinense* is sequenced and analyzed, which the size is 155,756 bp and with a GC content of 37.8%. This CP genome is made up of a large single copy region (LSC) and a small single copy region (SSC) with length of 86,595 and 18,209 bp, respectively. Also, two inverted repeat regions (IRa and IRb) with length of 25,476 bp were also a part of the CP genome, which were separated by the SCs. The genome encodes 114 genes, 16 of which are duplicated genes. Most of the 85 protein-coding genes(CDS) have a usual ATG start codon, except for 3 genes including rps12, psbL and ndhD. In addtion, a strong A/T bias was found in a majority of the simple sequence repeats(SSRs). Phylogenetic relationships among the 16 species reveals that *L. chinense* was sister to Atropa belladonna. In general, revealing of the complete sequences and annotation of the *L. chinense* chloroplast genome will provide valuable genetic information for identifying taxonomy, species, phylogenetic evolution of the family of Solanaceae precisely.

Introduction

Lycium chinense Mill. has an indisputable position at the magnificent Chinese traditional medicine history(Potterat 2010). It has been used not only as food and herbal medicine in China for thousands of years, but also as a functional food by more and more people all over the world. (Luo et al. 2006). Fruits of *L. chinense* possess potential pharmacological effects such as anti-aging, reducing blood glucose and serum lipids, immune regulation, etc(Gan et al. 2004, Qin et al. 2001). Moreover, the dry root bark of *L. chinense*, which is widely used for the treatment of night sweats, diabetes, coughs, vomiting blood, high blood pressure, and ulcers, is listed in the Chinese Pharmacopoeia (2015 version) officially .(Pharmacopoeia 2015).

L. chinense belongs to the family of Solanaceae. The Solanaceae family is composed of about 27 thousand species that belong to 24 genus, among which the Lycium is one of the most important genus. The Lycium genus contains approximately 70 species widely distributed throughout the world, including southern Africa, Europe, Asia, America and Australia(Turchetto et al. 2014, Zhang J. X. et al. 2013), while 7 of which are unique located in China. These 7 species are all deciduous shrubbery, possessing highly similar morphologies and structures, which makes it difficult to distinguish them by their appearance. As a result, they are often confused in the market(NI Lianghong 2016). DNA barcoding have been used to identify and to analysis the phylogenetic relationships of the Lycium genus. Yet, it is not effective to identify the genus using a few DNA barcode fragments. Therefore, it still needs more e□ective molecular markers to investigate the relationships within the Lycium genus(Hebert et al. 2003, NI Lianghong 2016)

In recent years, chloroplast (CP) genome have been widely used to reconstruct phylogenetic relationships among various land plants. As an important plastid, chloroplast plays an indispensable role in plant cell for photosynthesis and carbon fixation. With the rapid development of next-generation sequencing technologies, sequencing the entire chloroplast genome become a normal job for most laboratories(Nielsen et al. 2013). According to reported

researches, the structure of CP genomes in angiosperms are highly converted with a length of about 150 kb and are composed of IRa, IRb, LSC and SSC, and the two IRs are separated by the two SCs (Sanchez-Puerta and Abbona 2014, Yang Y. et al. 2014). Chloroplast genomes play a significant role in studying the evolutionary relationship at taxonomic level in plants as a result of being maternal inheritance, haploid, and high conservation in gene content and genome structure.(Shaw et al. 2007, Shaw et al. 2005).

The transcriptome analysis about the *L. chinense* leaf have been reported (Wang G. et al. 2015), however, information about its CP genomic structure is still unknown. Here in this article, in order to get a comprehensive understanding of the CP genome of *L. chinense*, the complete CP genome sequence is reported based on next-generation sequencing methods. Also, gene structure characteristics, RNA editing sites, codon usage as well as phylogenetic position of *L. chinense* are analyzed and compared with several related species. As to our knowledge, this is also the first comprehensive analysis on CP genome for the Lycium genus.

Materials & Methods

Plant material and DNA extraction

Fresh leaves of the *L. chinense* were obtained from the Medicinal Plant Garden of Guangzhou University of Chinese Medicine. Total genomic DNA (gDNA) was extracted from those leaves using a DNeasy Plant Mini Kit (Qiagen, German).

Chloroplast Genome Sequencing and Assembly

Sequencing library was constructed using this gDNA, after being ultrasonically sheered into 250 bp fragments, and then be submitted to Next-generation Sequencing on an Illumina HiSeq 2000 platform. NGS platform generated 6.82 G of raw sequencing data, which was then undergoes quality filtering and trimming to clean reads. Using the complete sequence of Atropa belladonna chloroplast genome as a reference, CP-like reads were extracted from those clean reads and then be assembled using the Abyss2.0 program(Jackman et al. 2017), resulting in a complete chloroplast genome sequence of *L. chinense*. PCR amplification was performed to verify the four junction regions between the IR regions and the LSC/SSC region.

Gene Annotation and Genome Structure

Gene annotation of the *L. chinense* CP genome was conducted using the online program GeSeq-Annotation of Organellar Genomes (https://chlorobox.mpimp-golm.mpg.de/geseq.html)(Tillich et al. 2017). Genious Pro. (version 4.8.4) was used to correct the annotation result by adjusting the open reading frame of coding genes. After that the complete chloroplast genome was submitted using the program Sequin and an accession number of MK040922 was assigned from the GenBank. The Organellar Genome DRAW (OGDRAW) program was used to draw the a physical map (Lohse et al. 2013).

GC content of CDS regions as well as the distribution of codon usage was analyzed using the Molecular Evolutionary Genetics Analysis (MEGA 6.06) (Kumar et al. 2008, Tamura et al.

2013).The online program Predictive RNA Editor for Plants (PREP) suite was used to predict potential RNA editing sites in 35 genes of the CP genome of *L. chinense* using (Mower 2009). Furthermore, by using MISA (Yang X. M. et al. 2012) , SSRs were detected and the forward and inverted repeats were determined with the REPuter program [https://bibiserv2.cebitec.uni-bielefeld.de/reputer] (Kurtz et al. 2001).

Genome Comparison and Phylogenetic Analysis

In order to compare the CP genome of *L. chinense* with others of Atropa belladonna, Capsicum annuum, Nicotiana tabacum and Datura stramonium, the mVISTA program(http://genome.lbl.gov/vista/index.shtml)(Mayor et al. 2000) in the Shuffle-LAGAN mode(Frazer et al. 2004) was carried out using the annotation of *L. chinense* as the reference.

In order to identify the phylogenetic position of *L. chinense* within the tubiflorae lineages, 16 complete CP genome sequences were downloaded from the GenBank of NCBI database. Those cp genomes underwent a sequence alignment by the MAFFT program(Katoh et al. 2017) (https://www.ebi.ac.uk/Tools/msa/muscle/). Maximum likelihood (ML) analysis of those 16 CP genomes was performed using Mega 6.06(Tamura et al. 2013) to find the best models and then construct ML tree, taking the CP genome sequences of Salvia japonica, Pogostemon cablin, and Andrographis paniculata as the outgroup. Support was estimated through 1,000 bootstrap replicates to assess the reliability of the phylogenetic tree.

Results and Discussion

CP Genome organization and gene content

The CP genome of *L. chinense* is 155,756 bp in length consisting of a large and small single-copy regions of length 86,595 and 18,209 bp, separated by two inverted repeat regions of 25,476 bp (Figure 1 and Table 1). The total G+C content is similar to that of other species in the Solanaceae family, which is about 37.8% (Amiryousefi A 2018, Cho et al. 2016, Sanchez-Puerta and Abbona 2014, Yang Y. et al. 2014). And the G+C contents of the LSC(35.8%) and SSC regions (32.3%)are lower than those of the IR regions (43.1%). As for the protein-coding regions (CDS), G+C content of the 1st , 2nd and 3rd codon positions are 43.9%, 37.9%, and 33.1%, respectively (Table 1), which means the A+T content is much higher at the third codon position than the other two. This phenomenon, which has been commonly found in other plant CP genomes, is used to separate the nuclear and mitochondrial DNA from CP DNA (Clegg et al. 1994, He et al. 2017, Morton 1998, Xiang et al. 2016).

It is predicted to encode a total of 130 predicted functional genes in the CP genome of *L. chinense*, while 16 of them, including 5 CDS, 7 tRNA genes and 4 rRNA genes, are duplicated in the IR region. The annotation revealed 80 distinct protein-coding genes, 30 distinct tRNA genes and 4 distinct rRNA genes (Table 2). Interestingly, there are 82 protein-coding genes have a regular ATG start codon within the 85 protein-coding genes, except for 3 genes, including ACG for psbL and ndhD, ACT for rps12. As start codons, ACG and ACT are meaningless, while as non-starting codons, they are synonymous codon which can still encode threonine. This

phenomenon has also appeared in model plant Nicotiana tabacum, in which the start codon for psbL and ndhD are also ACG(Kahlau et al. 2006). Furthermore, ACT start codon for rps12 can also be found in mitochondrial genome of tube-dwelling diatom Berkeleya fennica(An et al. 2016). Therefore, we hypothesis that the start codon of some genes such as rps12, psbL and ndhD may be mutated during the evolution. In addition, one gene (ycf1)without a stop codon were annotated as a pseudogene. Intron-containing genes were also analysis in this article. In total, there are 17 genes which contains one or two introns found in the CP genomes of *L. chinense* (Table 3), including 12 CDS and 5 tRNA genes, as listed in the table, while two of them, ycf3 and clpP, contain two intron and a single intron was detected in the other 15 genes, which has also been reported in other plants (Guo et al. 2017, Shen et al. 2018) .

Repeat Structure and SSR Analysis

SSRs, also called microsatellites, which are widely distributed across the entire genome, refer to a group of tandem sequences. (Chen et al. 2006). SSRs play a key role in a genome and have been widely used in genetic and genomic studies as a result of their extreme variability within species (Huang et al. 2014, Vieira Ldo et al. 2014, Wheeler et al. 2014, Zhao et al. 2014).The SSRs of *L. chinense* were detected and presented in Table 4. There are 107 mononucleotide, 46 dinucleotide, 67 trinucleotide and 11 tetranucleotide repeat units detected, making a total of 231 SSR loci. In addition, 99.1% of the mononucleotide SSRs were constituted by A/T sequences, while only one is composed of a G/C motif. Interestingly, 63.0% of the dinucleotide SSRs were also constituted of A/T motifs. These results are all consistent with the hypothesis that CP SSRs have a strong A/T bias, which is found to be common in many other plants (Firetti et al. 2017, Wang W. et al. 2018, Zhou et al. 2018)The abundant AT in CP genome may be related to this, which is related to the stability of AT and GC to some extent(Yang Y. et al. 2014).

Long repeats are sequence repeats of length equal or greater than 30 bp, which may have functions to increase population genetic diversity and promote chloroplast genome rearrangement (Qian et al. 2013). It must be noted that the repeat types found are all forward and palindromic among the four species. There are 49 (25 forward, 24 palindromic), 40 (20 forward, 20 palindromic), 42 (25 forward, 17 palindromic), 48 (25 forward, 23 palindromic) large repeats in the CP genomes of *L. chinense*, Atropa belladonna, Capsicum annuum, Nicotiana tabacum, respectively(Figure 2).

Comparative Chloroplast Genomic Analysis

In order to proceed the subsequent phylogenetic analyses and plant identification smoothly, mVISTA program was carried out to analyze the whole CP genome sequence of *L. chinense* and was compared to that of Atropa belladonna (NC_004561.1), Capsicum annuum(NC_018552.1), Nicotiana tabacum(Z00044.2) and Datura stramonium(NC_018117.1) (Figure 3). It can be seen clearly from the figure that the IR regions are more conservative than the SC regions. Copy correction caused by gene conversion between the two IR region sequences may be the main cause of this phenomenon (Khakhlova and Bock 2006). Moreover, higher conservatism was

observed in the coding regions than in the non-coding regions, which is very common in the CP genomes of many other angiosperms(Chen et al. 2017, Cheng et al. 2017, Kong and Yang 2017).

IR Contraction and Expansion in the *L.chinense* CP Genome

As shown in Figure 4， the IR-SSC and IR-LSC boundaries of *L. chinense*. were compared to that of three other species, including Atropa belladonna (NC_004561.1), Capsicum annuum (NC_018552.1) and Solanum lycopersicum (NC_007898.3). All of the four species belong to the Solanaceae family. The length of the IR region in the four CP genomes ranged from 25,476 bp to 25,906 bp, showing a modest expansion. Due to the expantion, the rps19 and the ycf1 gene was partially included in the IR regions of the Solanoideae family. As a result, there is a truncated rps19 pseudogene and a ycf1 pseudogene copy found at the junction of LSC/IRB and SSC/IRA, respectively. The ndhF gene, which is located in the SSC region entirely, showing a variety distance from the LSC edge within the Solanoideae family, while the longest distance(54 bp） is observed in the the *L. chinense*. On the other hand, 47 bp of the rps19 gene and 995 bp of the ycf1 gene are extended into the IR regions in the *L. chinense*. CP genome, which is the shortest among the four species. In the meantime, in A.belladonna, C.annuumand and S.lycopersicum, 60, 66 and 91 bp of the rps19 gene and 1,438, 1,128 and 1,119 bp of the ycf1 gene are extended into the IR regions, respectively. Furthermore, the length of the IR region (25476 bp) in *L. chinense*. CP genome is also the shortest among the four species and this might be due to the shortest IR expansion of the rps19 gene and the ycf1 gene. Interestingly, the size of the *L. chinense*. CP genome(155756bp) is also the shortest, which is consistent with the hypothesis that the IR contraction and expansion is the main reason to explain size differences between CP genomes(Li et al. 2013, Wang R. J. et al. 2008). Taken together, we can see that the expansion and contraction of IR/SC regions have a similar pattern within family though it still varied slightly.

Codon usage and RNA Editing Sites

As for codon usage analysis, the results were summarized in Figure 5 and Table S1. We found that there are 20 amino acids which can be transported for protein biosynthesis by the tRNA found in the *L. chinense* CP Genome. Moreover, all the CDS were consisted of 26,569 codons, among which the usage of codons encoding leucine was the highest, accounting for 13.16% of the total usage, while the usage of codons encoding cysteine was the lowest, accounting for 1.82% of the total usage in the *L. chinense* CP Genome. Furthermore, as the number of codons encoding a particular amino acid increases, the value of RSCU (shortening of relative synonymous codon usage) also increases, as Figure 5 shows. Interestingly, most of the amino acid codons, in addition to two of them, which are methionine and tryptophan, have preferences, and is the same as the other species(Chang et al. 2006, Pan et al. 2012, Reginato et al. 2016, Wu et al. 2010, Zhang Y. et al. 2016).

As a common phenomenon in plant CP genomes, RNA editing participate in the process of modifing mutations, changing reading frames, and regulating the expression of chloroplast genes (Freyer et al. 1993). The program PREP(Predictive RNA Editor for Plants) suite was used to

observe the potential RNA editing sites of the *L. chinense* CP genome, and the results were analyzed and summarized in Table S2. From the table we can see that a total of 50 RNA editing sites in 35 genes were identified and all the nucleotide changes found are cytidine(C)-Thymine(T) editing, which are always happened in the transcripts of land plant chloroplast genomes (Tsudzuki et al. 2001). Due to the RNA editing，the ndhD and psbL can be normally transcribed, and this phenomenon are also found in the Linum usitatissimum L(de Santana Lopes et al. 2018), tobacco (Hirose and Sugiura 1997),spinach(Maier et al. 1996), and Ampelopsis brevipedunculata(Raman and Park 2016). Furthermore, the results also reveals that the conversion of amino acid from S to L has the highest frequency of occurrence, which accounted for about 40%; while H-Y and I-F has the lowest frequency of occurrence, accounting for only 2%.

Phylogenetic Analysis

With the great potential in studies of phylogenetics, evolution and molecular systematic, chloroplast genomes has been widely used to solve the phylogenetic questions in many land plants(Amiryousefi A 2018, De Las Rivas et al. 2002, Zhou et al. 2018). For purpose of identifying the evolutionary position of *L. chinense* within the Solanaceae family, multiple sequence alignments using 12 other Solanaceae partial chloroplast genome sequences, which only contain one IR region, were carried out. Three species from different family in Tubiflorae were also chosen as outgroups (Table S3). The result was shown in Figure 6, in which most nodes were strongly supported by 100 % bootstrap values. Furthermore, all of the 16 species splits into two clades, while the 13 Solanaceae plants composed a unique clade, and the three outgroup species gather into the other clade. As for Solanaceae, different genus basically can be clustered alone, which demonstrates good monophylaxis among this family. And *L. chinense* (Lycium) was sister to Atropa belladonna (Atropa). Overall, this study will promote the use of chloroplast genome for species identification.

Conclusions

The complete CP genome of *L. chinense* is first reported in this article. And it is also the first species within the Lycium genus to have the CP genome fully sequenced and analysis. The CP genome of *L. chinense* is 155,756 bp in length and has the relatively conservative genome structure as well as gene content. Forty-nine repeated sequences and 231 SSRs, which are informative sources for the development of new molecular makers, were determined and analyzed. The genome structure and composition among the four species are similar. However, the CP genome of *L. chinense* has the shortest size compared to other three species, which might be caused by the IR contraction. By comparing the CP genome organization of different species, it will help us understand the evolution process of chloroplast more deeply. The results of the phylogenetic analysis, which is conducted among the 16 species, demonstrate that *L. chinense* has a close relationship with Atropa belladonna. In a word, the data, which has been published in this paper, will promote the further investigation of *L. chinense*.

Acknowledgements

no.

References

Amiryousefi A HJ, Poczai P. 2018. The chloroplast genome sequence of bittersweet (Solanum dulcamara): Plastid genome structure evolution in Solanaceae. *PLoS One* 13: 23. DOI 10.1371/journal.pone.0196069.g001

An SM, Noh JH, Choi DH, Lee JH, Yang EC. 2016. Repeat region absent in mitochondrial genome of tube-dwelling diatom Berkeleya fennica (Naviculales, Bacillariophyceae). *Mitochondrial DNA A DNA Mapp Seq Anal* 27: 2137-2138. DOI 10.3109/19401736.2014.982594

Chang CC, Lin, H. C, Lin, I. P, Chow, T. Y, Chen, H. H, Chen, W. H, Cheng, C. H, Lin, C. Y, Liu, S. M, Chang, C. C, Chaw, S. M. 2006. The chloroplast genome of Phalaenopsis aphrodite (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol Biol Evol* 23: 279-291. DOI 10.1093/molbev/msj029

Chen C, Zhou P, Choi YA, Huang S, Gmitter FG, Jr. 2006. Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet* 112: 1248-1257. DOI 10.1007/s00122-006-0226-1

Chen C, Zheng Y, Liu S, Zhong Y, Wu Y, Li J, Xu LA, Xu M. 2017. The complete chloroplast genome of Cinnamomum camphora and its comparison with related Lauraceae species. *PeerJ* 5: e3820. DOI 10.7717/peerj.3820

Cheng H, Li J, Zhang H, Cai B, Gao Z, Qiao Y, Mi L. 2017. The complete chloroplast genome sequence of strawberry (Fragaria x ananassa Duch.) and comparison with related species of Rosaceae. *PeerJ* 5: e3919. DOI 10.7717/peerj.3919

Cho KS, Cheon KS, Hong SY, Cho JH, Im JS, Mekapogu M, Yu YS, Park TH. 2016. Complete chloroplast genome sequences of Solanum commersonii and its application to chloroplast genotype in somatic hybrids with Solanum tuberosum. *Plant Cell Rep* 35: 2113-2123. DOI 10.1007/s00299-016-2022-y

Clegg MT, Gaut BS, Learn GH, Jr., Morton BR. 1994. Rates and patterns of chloroplast DNA evolution. *Proc Natl Acad Sci USA*, 91: 6795-6801.

De Las Rivas J, Lozano JJ, Ortiz AR. 2002. Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Res* 12: 567-583. DOI 10.1101/gr.209402

Amanda de Santana Lopes A, Pacheco TG, Santos KGD, Vieira LDN, Guerra MP, Nodari RO, de Souza EM, de Oliveira Pedrosa F, Rogalski M. 2018. The *Linum usitatissimum* L. plastome reveals atypical structural evolution, new editing sites, and the phylogenetic position of Linaceae within Malpighiales. *Plant Cell Rep* 37: 307-328. DOI 10.1007/s00299-017-2231-z

Firetti F, Zuntini AR, Gaiarsa JW, Oliveira RS, Lohmann LG, Van Sluys MA. 2017. Complete chloroplast genome sequences contribute to plant species delimitation: A case study of the Anemopaegma species complex. *Am J Bot* 104: 1493-1509. DOI 10.3732/ajb.1700302

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* 32: W273-279. DOI 10.1093/nar/gkh458

Freyer R, Hoch B, Neckermann K, Maier RM, Kossel H. 1993. RNA editing in maize chloroplasts is a processing step independent of splicing and cleavage to monocistronic mRNAs. *Plant J* 4: 621-629.

Gan L, Hua Zhang S, Liang Yang X, Bi Xu H. 2004. Immunomodulation and antitumor activity by a polysaccharide-protein complex from Lycium barbarum. *Int Immunopharmacol* 4: 563-569. DOI 10.1016/j.intimp.2004.01.023

Guo H, Liu J, Luo L, Wei X, Zhang J, Qi Y, Zhang B, Liu H, Xiao P. 2017. Complete chloroplast genome sequences of Schisandra chinensis: genome structure, comparative analysis, and phylogenetic relationship of basal angiosperms. *Sci China Life Sci* 60: 1286-1290. DOI 10.1007/s11427-017-9098-5

He L, Qian J, Li X, Sun Z, Xu X, Chen S. 2017. Complete Chloroplast Genome of Medicinal Plant Lonicera japonica: Genome Rearrangement, Intron Gain and Loss, and Implications for Phylogenetic Studies. *Molecules* 22. DOI 10.3390/molecules22020249

Hebert PD, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. *Proc Biol Sci* 270: 313-321. DOI 10.1098/rspb.2002.2218

Hirose T, Sugiura M. 1997. Both RNA editing and RNA cleavage are required for translation of tobacco chloroplast ndhD mRNA: a possible regulatory mechanism for the expression of a chloroplast operon consisting of functionally unrelated genes. *EMBO J* 16: 6804-6811. DOI 10.1093/emboj/16.22.6804

Huang H, Shi C, Liu Y, Mao SY, Gao LZ. 2014. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. *BMC Evol Biol* 14: 151. DOI 10.1186/1471-2148-14-151

Jackman SD, Jackman, S. D, Vandervalk, B. P, Mohamadi, H, Chu, J, Yeo, S, Hammond, S. A, Jahesh, G, Khan, H, Coombe, L, Warren, R. L, Birol, I. 2017. ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter. *Genome Res* 27: 768-777. DOI 10.1101/gr.214346.116

Kahlau S, Aspinall S, Gray JC, Bock R. 2006. Sequence of the tomato chloroplast DNA and evolutionary comparison of solanaceous plastid genomes. *J Mol Evol* 63: 194-207. DOI 10.1007/s00239-005-0254-5

Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. DOI 10.1093/bib/bbx108

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J* 46: 85-94. DOI 10.1111/j.1365-313X.2006.02673.x

Kong WQ, Yang JH. 2017. The complete chloroplast genome sequence of Morus cathayana and Morus multicaulis, and comparative analysis within genus Morus L. *PeerJ* 5: e3037. DOI 10.7717/peerj.3037

Kumar S, Nei M, Dudley J, Tamura K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* 9: 299-306. DOI 10.1093/bib/bbn017

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29: 4633-4642.

Li R, Ma PF, Wen J, Yi TS. 2013. Complete sequencing of five araliaceae chloroplast genomes and the phylogenetic implications. *PLoS One* 8: e78568. DOI 10.1371/journal.pone.0078568

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: W575-581. DOI 10.1093/nar/gkt289

Luo Q, Li Z, Huang X, Yan J, Zhang S, Cai YZ. 2006. Lycium barbarum polysaccharides: Protective effects against heat-induced damage of rat testes and H2O2-induced DNA damage in mouse testicular cells and beneficial effect on sexual behavior and reproductive function of hemicastrated rats. *Life Sci* 79: 613-621. DOI 10.1016/j.lfs.2006.02.012

Maier RM, Zeltz P, Kossel H, Bonnard G, Gualberto JM, Grienenberger JM. 1996. RNA editing in plant mitochondria and chloroplasts. *Plant Mol Biol* 32: 343-365.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. 2000. VISTA : visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* 16: 1046-1047.

Morton BR. 1998. Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. *J Mol Evol* 46: 449-459.

Mower JP. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res* 37: W253-259. DOI 10.1093/nar/gkp337

NI Lianghong ZZ, LU Jiani 2016. DNA barcoding construction of medicinal plants in genus Lycium L. based on multiple genomic segments. *Chinese Traditional and Herbal Drugs* 47: 5.

Nielsen AZ, Ziersen B, Jensen K, Lassen LM, Olsen CE, Moller BL, Jensen PE. 2013. Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth Biol* 2: 308-315. DOI 10.1021/sb300128r

Pan IC, Liao DC, Wu FH, Daniell H, Singh ND, Chang C, Shih MC, Chan MT, Lin CS. 2012. Complete chloroplast genome sequence of an orchid model plant candidate: Erycina pusilla apply in tropical Oncidium breeding. *PLoS One* 7: e34738. DOI 10.1371/journal.pone.0034738

Pharmacopoeia CoN. 2015. China pharmacopoeia.

Potterat O. 2010. Goji (Lycium barbarum and *L. chinense*): Phytochemistry, pharmacology and safety in the perspective of traditional uses and recent popularity. *Planta Med* 76: 7-19. DOI 10.1055/s-0029-1186218

Qian, J, Song, J, Gao, H, Zhu, Y, Xu, J, Pang, X, Yao, H, Sun, C, Li, X, Li, C, Liu, J, Xu, H, Chen, S. et al. 2013. The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. *PLoS One* 8: e57607. DOI 10.1371/journal.pone.0057607

Qin X, Yamauchi R, Aizawa K, Inakuma T, Kato K. 2001. Structural features of arabinogalactan-proteins from the fruit of Lycium chinense Mill. *Carbohydr Res* 333: 79-85.

Raman G, Park S. 2016. The Complete Chloroplast Genome Sequence of Ampelopsis: Gene Organization, Comparative Analysis, and Phylogenetic Relationships to Other Angiosperms. *Front Plant Sci* 7: 341. DOI 10.3389/fpls.2016.00341

Reginato M, Neubig KM, Majure LC, Michelangeli FA. 2016. The first complete plastid genomes of Melastomataceae are highly structurally conserved. *PeerJ* 4: e2715. DOI 10.7717/peerj.2715

Sanchez-Puerta MV, Abbona CC. 2014. The chloroplast genome of Hyoscyamus niger and a phylogenetic study of the tribe Hyoscyameae (Solanaceae). *PLoS One* 9: e98353. DOI 10.1371/journal.pone.0098353

Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot* 94: 275-288. DOI 10.3732/ajb.94.3.275

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am J Bot* 92: 142-166. DOI 10.3732/ajb.92.1.142

Shen X, et al. 2018. Complete Chloroplast Genome Sequence and Phylogenetic Analysis of Aster tataricus. *Molecules* 23. DOI 10.3390/molecules23102426

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30: 2725-2729. DOI 10.1093/molbev/mst197

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

Tsudzuki T, Wakasugi T, Sugiura M. 2001. Comparative analysis of RNA editing sites in higher plant chloroplasts. *J Mol Evol* 53: 327-332. DOI 10.1007/s002390010222

Turchetto C, Fagundes NJ, Segatto AL, Kuhlemeier C, Solis Neffa VG, Speranza PR, Bonatto SL, Freitas LB. 2014. Diversification in the South American Pampas: the genetic and morphological variation of the widespread Petunia axillaris complex (Solanaceae). *Mol Ecol* 23: 374-389. DOI 10.1111/mec.12632

Vieira Ldo N, Faoro H, Rogalski M, Fraga HP, Cardoso RL, de Souza EM, de Oliveira Pedrosa F, Nodari RO, Guerra MP. 2014. The complete chloroplast genome sequence of

Podocarpus lambertii: genome structure, evolutionary aspects, gene content and SSR detection. *PLoS One* 9: e90618. DOI 10.1371/journal.pone.0090618

Wang G, Du X, Ji J, Guan C, Li Z, Josine TL. 2015. De novo characterization of the Lycium chinense Mill. leaf transcriptome and analysis of candidate genes involved in carotenoid biosynthesis. *Gene* 555: 458-463. DOI 10.1016/j.gene.2014.10.058

Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM. 2008. Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evol Biol* 8: 36. DOI 10.1186/1471-2148-8-36

Wang W, Chen S, Zhang X. 2018. Whole-Genome Comparison Reveals Divergent IR Borders and Mutation Hotspots in Chloroplast Genomes of Herbaceous Bamboos (Bambusoideae: Olyreae). *Molecules* 23. DOI 10.3390/molecules23071537

Wheeler GL, Dorman HE, Buchanan A, Challagundla L, Wallace LE. 2014. A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Appl Plant Sci* 2. DOI 10.3732/apps.1400059

Wu FH, Chan MT, Liao DC, Hsu CT, Lee YW, Daniell H, Duvall MR, Lin CS. 2010. Complete chloroplast genome of Oncidium Gower Ramsey and evaluation of molecular markers for identification and breeding in Oncidiinae. *BMC Plant Biol* 10: 68. DOI 10.1186/1471-2229-10-68

Xiang B, Li X, Qian J, Wang L, Ma L, Tian X, Wang Y. 2016. The Complete Chloroplast Genome Sequence of the Medicinal Plant Swertia mussotii Using the PacBio RS II Platform. *Molecules* 21. DOI 10.3390/molecules21081029

Yang XM, Sun JT, Xue XF, Zhu WC, Hong XY. 2012. Development and characterization of 18 novel EST-SSRs from the western flower Thrips, Frankliniella occidentalis (Pergande). *Int J Mol Sci* 13: 2863-2876. DOI 10.3390/ijms13032863

Yang Y, Dang Y, Li Q, Lu J, Li X, Wang Y. 2014. Complete chloroplast genome sequence of poisonous and medicinal plant Datura stramonium: organizations and implications for genetic engineering. *PLoS One* 9: e110656. DOI 10.1371/journal.pone.0110656

Zhang JX, Guan SH, Feng RH, Wang Y, Wu ZY, Zhang YB, Chen XH, Bi KS, Guo DA. 2013. Neolignanamides, lignanamides, and other phenolic compounds from the root bark of Lycium chinense. *J Nat Prod* 76: 51-58. DOI 10.1021/np300655y

Zhang Y, et al. 2016. The Complete Chloroplast Genome Sequences of Five Epimedium Species: Lights into Phylogenetic and Taxonomic Analyses. *Front Plant Sci* 7: 306. 10.3389/fpls.2016.00306

Zhao Y, et al. 2014. The complete chloroplast genome provides insight into the evolution and polymorphism of Panax ginseng. *Front Plant Sci* 5: 696. 10.3389/fpls.2014.00696

Zhou T, Wang J, Jia Y, Li W, Xu F, Wang X. 2018. Comparative Chloroplast Genome Analyses of Species in Gentiana section Cruciata (Gentianaceae) and the Development of Authentication Markers. *Int J Mol Sci* 19. DOI 10.3390/ijms19071962

Figure 1

Map of *L. chinense* Mill plastid.

Genes inside the circle are transcribed clockwise, while genes ousideside the circle counterclockwise transcribed. Genes in different functional groups are color coded following the legend. The gray arrow represents gene direction. The darker gray in the inner circle denotes to GC content, whereas the lighter gray denotes to AT content.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27305v1 | CC BY 4.0 Open Access | rec: 30 Oct 2018, publ: 30 Oct 2018

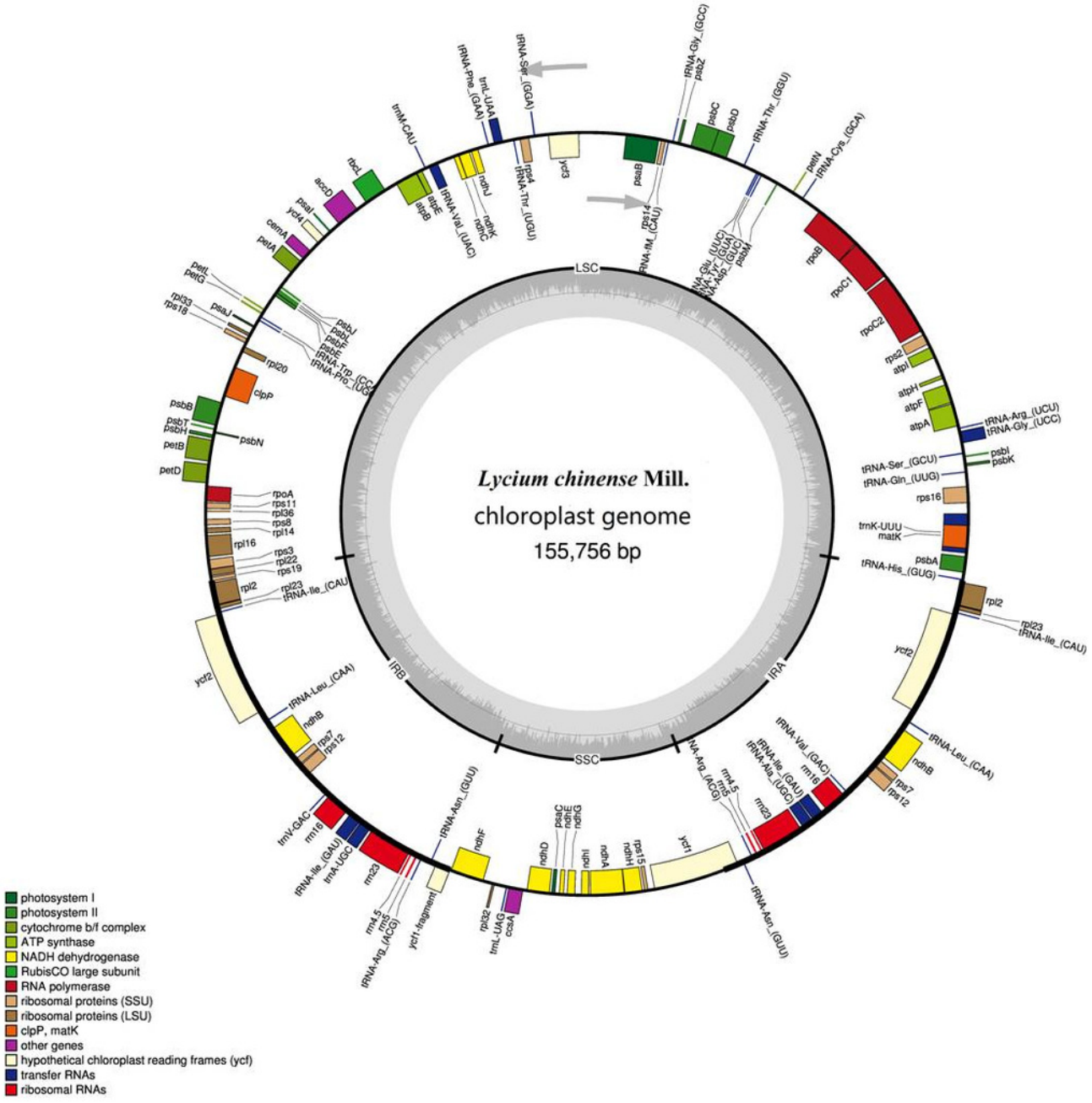

Figure 2

Long repeat sequences among the four chloroplast genomes.

F, P, R, and C indicates the repeat types F (forward), P (palindrome), R (reverse), and C (complement), respectively. Different colours represent repeats with different lengths.

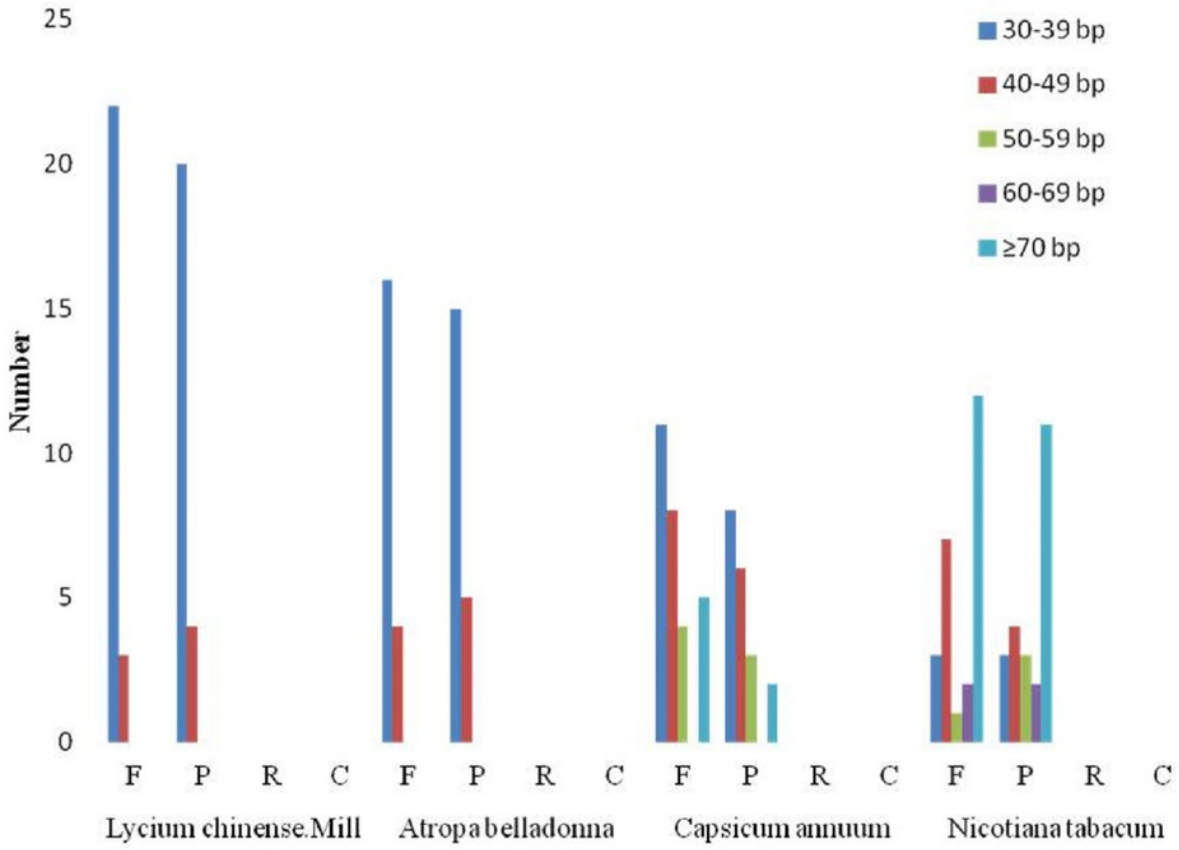

Figure 3

Sequence identity plot comparing the five chloroplast genomes by using mVISTA with *L. chinense* as a reference.

Grey arrows reveals the direction and position of each genes.The Y-scale represents the percent identity ranging from 50% to 100%.

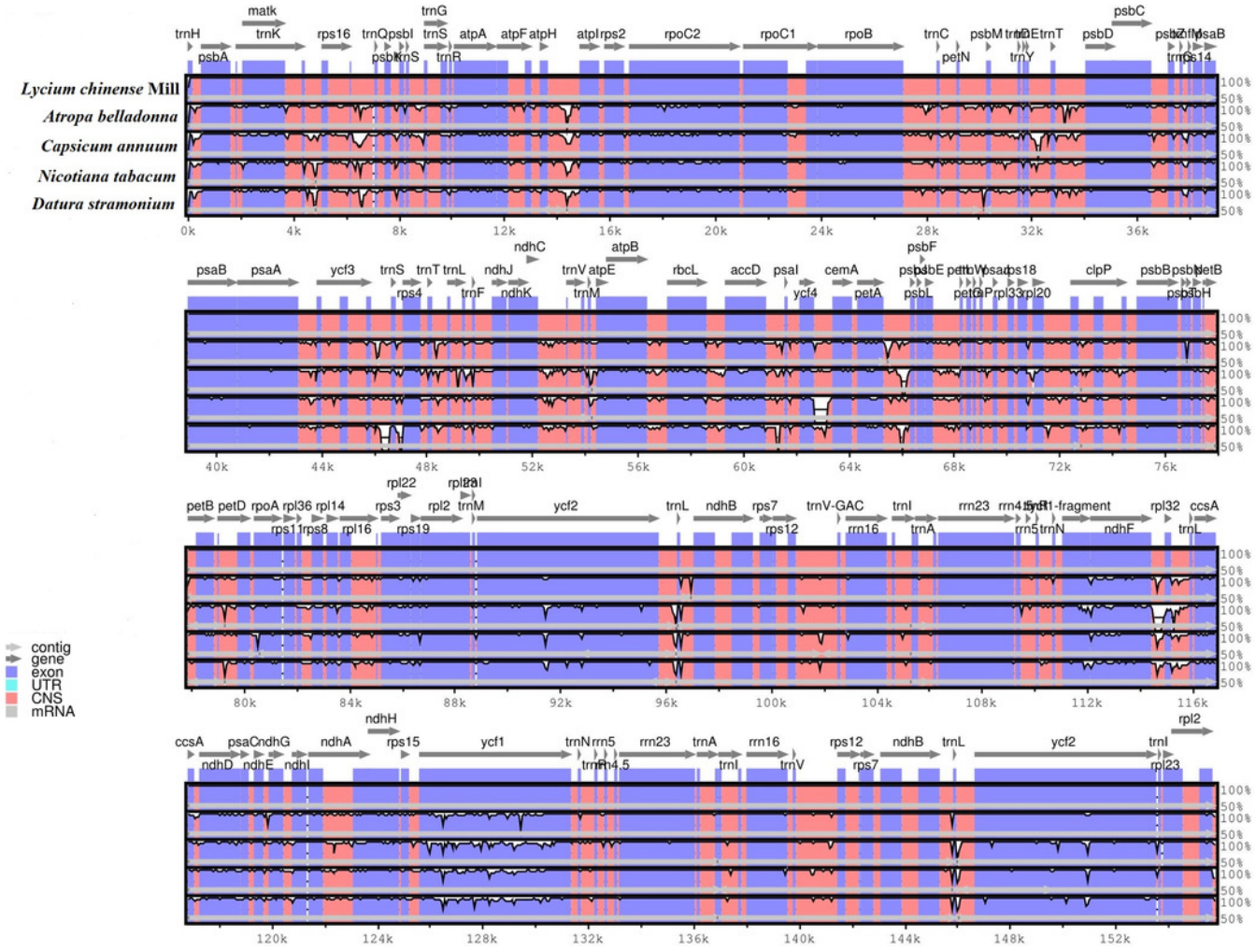

Figure 4

Comparison of the borders of LSC and SSC and IR regions among four chloroplast genomes

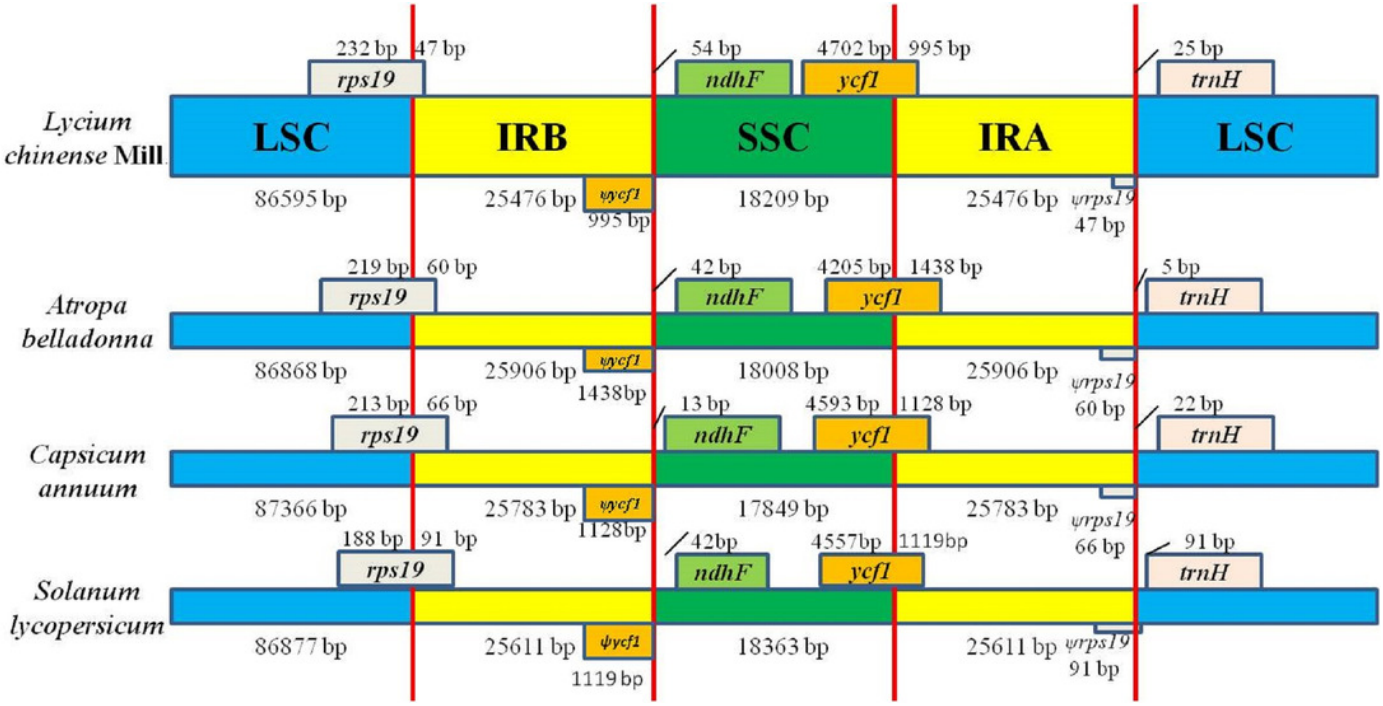

Figure 5

Codon content of 20 amino acid and stop codons in all protein-coding genes of the chloroplast of *L. chinense*

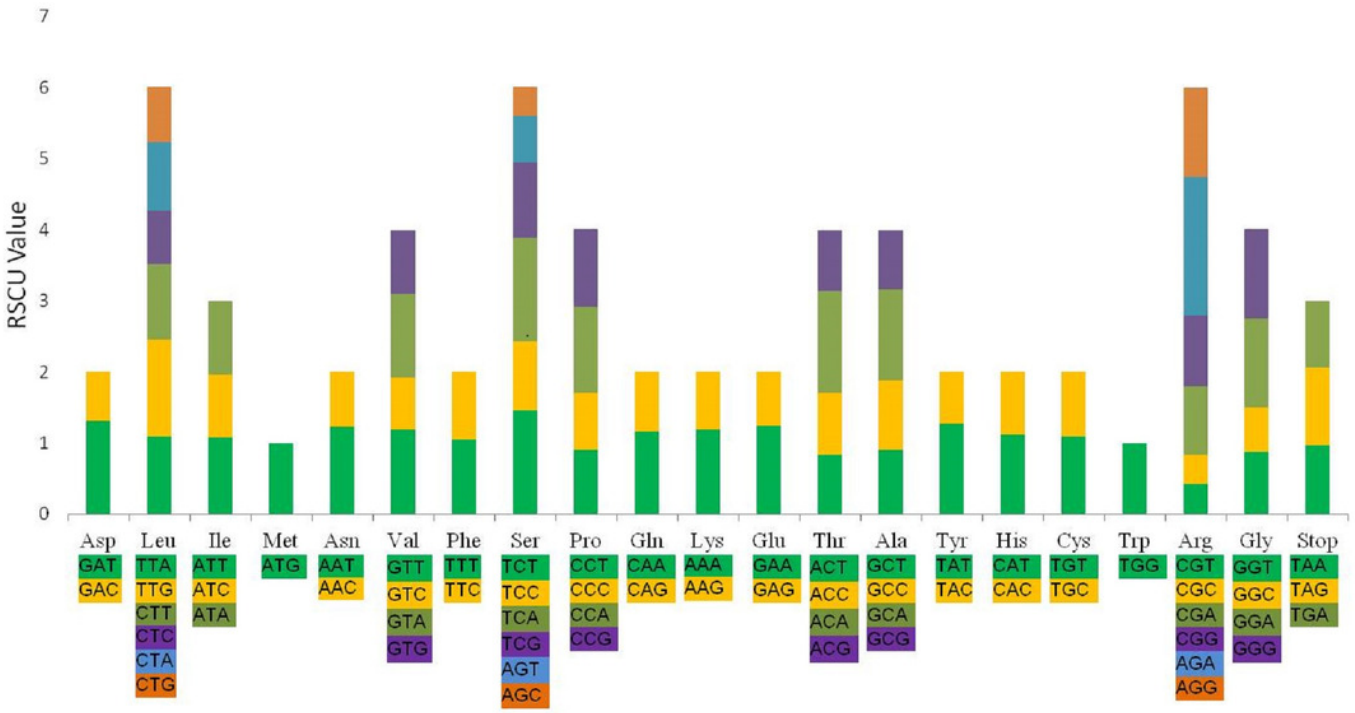

Figure 6

Phylegenetic relationshipp of the 16 species inferred from ML analysis based on the complete chloroplast genome without IRA region

Numbers at nodes are values for bootstrap support. The position of *L. chinense* is indicated in block letter. *Salvia japonica* , *Pogostemon cablin* and *Andrographis paniculata* are set as outgroup.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27305v1 | CC BY 4.0 Open Access | rec: 30 Oct 2018, publ: 30 Oct 2018

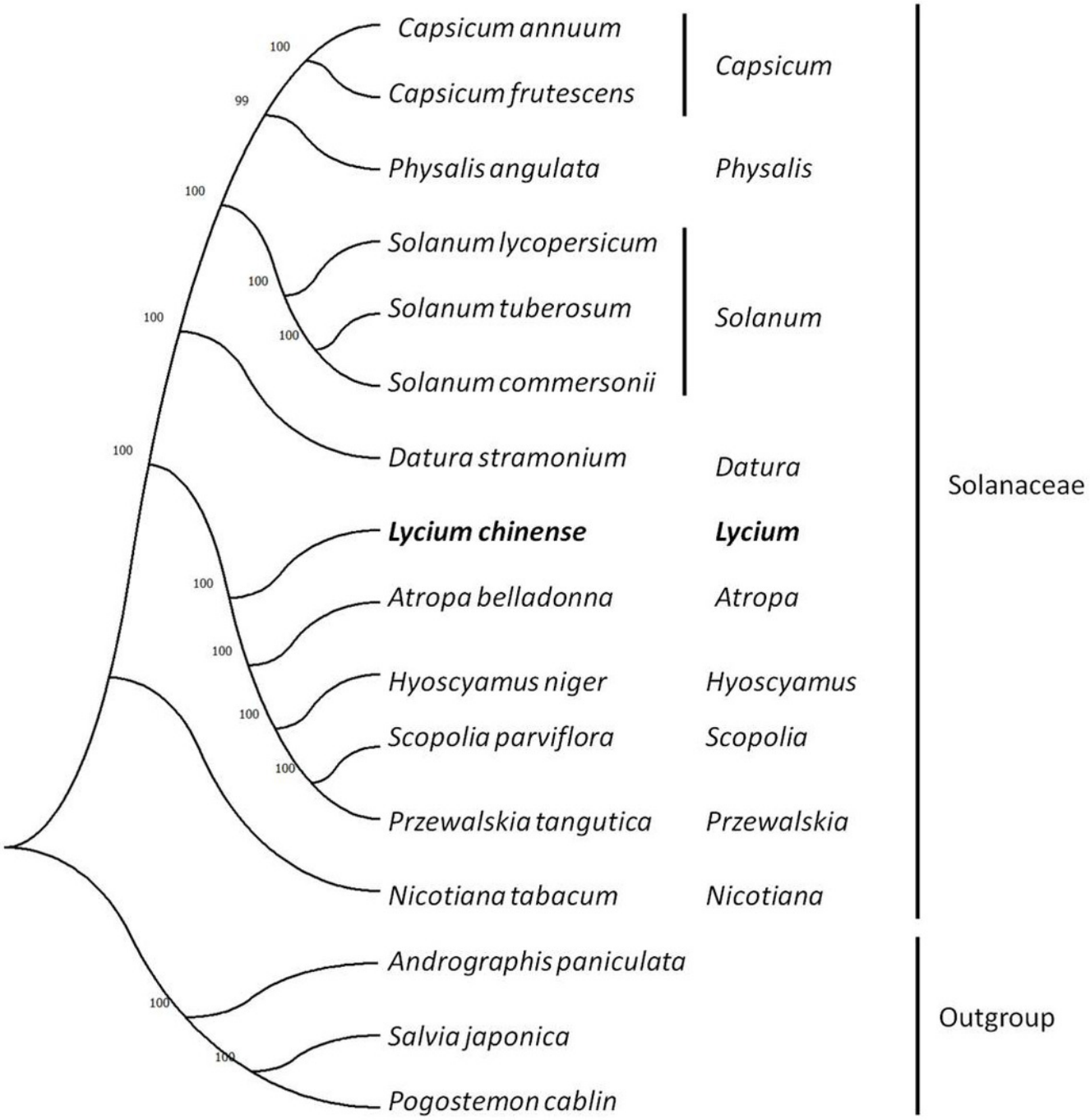

Table 1(on next page)

Base composition in the chloroplast genome of *L. chinense*

1

Region	Positions	T/U(%)	C(%)	A(%)	G(%)	Total(bp)
LSC		32.7	18.3	31.4	17.5	86,595
IRB		28.4	20.7	28.5	22.4	25,476
SSC		33.9	16.8	33.7	15.5	18,209
IRA		28.5	22.4	28.4	20.7	25,476
Total		31.5	19.2	30.7	18.6	155,756
CDS		31.3	17.9	30.5	20.3	79,700
	1st position	25.0	18.6	30.9	25.3	26,567
	2nd position	34.0	19.6	28.5	18.3	26,567
	3rd position	35.0	15.5	32.7	17.6	26,566

2

Table 2(on next page)

Gene contents in the chloroplast genomes of *L. chinense*

1

Classification of Genes	Gene Names	Amount
Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*	5
Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*	15
Cytochrome b/f complex	*petA*, *petB**, *petD**, *petG*, *petL*, *petN*	6
ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*	6
NADH dehydrogenase	*ndhA**, *ndhB**(×2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*	12（1）
RubisCO large subunit	*rbcL*	1
RNA polymerase	*rpoA*, *rpoB*, *rpoC1*, *rpoC2*	4
Ribosomal proteins (SSC)	*rps2*, *rps3*, *rps4*, *rps7* (×2), *rps8*, *rps11*, *rps12*** (×2), *rps14*, *rps15*, *rps16**, *rps18*, *rps19*	14（2）
Ribosomal proteins (LSC)	*rpl2* (×2), *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23* (×2), *rpl32*, *rpl33*, *rpl36*	11
Ribosomal RNAs	*rrn 4.5* (×2), *rrn 5* (×2), *rrn 16* (×2), *rrn 23* (×2)	8（4）
Protein of unkown function	*ycf1* (×2), *ycf2* (×2), *ycf3***, *ycf4*	6（2）
Transfer RNAs	37 tRNAs (8 contain an intron, 7 in the inverted repeats region)	37（7）
Other genes	*accD*, *ccsA*, *cemA*, *clpP*, *matK*	5
Total		130

2

Table 3(on next page)

Gene with introns in the chloroplast genomes of *L. chinense* as well as the exons

Gene	Location	Exon Ⅰ (bp)	Intron Ⅰ (bp)	Exon Ⅱ (bp)	Intron Ⅱ (bp)	Exon Ⅲ (bp)
atpF	LSC	410	704	145		
clpP	LSC	228	640	292	808	71
ndhA	SSC	539	1154	553		
ndhB	IR	777	679	756		
petB	LSC	6	750	642		
petD	LSC	8	742	475		
rpl16	LSC	396	1016	9		
rpl2	IR	434	666	391		
rpoC1	LSC	1616	737	430		
rps12	IR	26	536	232		
rps16	LSC	227	822	40		
trnA-UGC	IR	38	681	35		
trnI-GAU	IR	34	717	37		
trnK-UUU	LSC	36	2513	37		
trnL-UAA	LSC	35	497	50		
trnV-UAC	LSC	35	565	38		
ycf3	LSC	154	756	229	744	124

1

Table 4(on next page)

Types and amounts of SSRs in the *L. chinense* chloroplast genomes

SSR Type	Repeat Unit	Amount	Ratio（%）
Mono	A/T	106	99.1
	C/G	1	0.9
Di	AC/GT	1	2.2
	AG/CT	16	34.8
	AT/TA	29	63.0
Tri	AAC/GTT	9	13.4
	AAG/CTT	20	30.0
	AAT/ATT	21	31.3
	ACC/GGT	1	1.4
	ACG/CGT	1	1.4
	ACT/AGT	2	3.0
	AGC/CTG	5	7.5
	AGG/CCT	4	6.0
	ATC/ATG	4	6.0
Tetra	AAAC/GTTT	3	27.3
	AAAG/CTTT	1	9.1
	AAAT/ATTT	5	45.4
	AATC/ATTG	1	9.1
	AGAT/ATCT	1	9.1

1