MITOCHONDRIAL DNA PART B
2023, VOL. 8, NO. 6, 643–647
https://doi.org/10.1080/23802359.2023.2220436

Taylor & Francis
Taylor & Francis Group

PLASTOME ANNOUNCEMENT

OPEN ACCESS Check for updates

Characterization of the complete chloroplast genome of *Gypsophila huashanensis* Y. W. Tsui & D. Q. Lu, an endemic herb species in China

Tian-Xia Guan[a], Zhao-Ping Lu[a], Mi-Li Liu[a], Lu-Lu Xun[b], Min-Feng Fang[a] and Zhong-Hu Li[a]

[a]Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an, Shaanxi, China; [b]Shaanxi Engineering Research Centre for Conservation and Utilization of Botanical Resources, Xi'an Botanical Garden of Shaanxi Province (Institute of Botany of Shaanxi Province), Xi'an, Shaanxi, China

ABSTRACT

Gypsophila huashanensis Y. W. Tsui & D. Q. Lu (Caryophyllaceae) is an endemic herb species to the Qinling Mountains in China. In this study, we characterized its whole plastid genome using the Illumina sequencing platform. The complete plastid genome of *G. huashanensis* is 152,457 bp in length, including a large single-copy DNA region of 83,476 bp, a small single-copy DNA region of 17,345 bp, and a pair of inverted repeat DNA sequences of 25,818 bp. The genome contains 130 genes comprising 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Evolutionary analysis showed that the non-coding regions of Caryophyllaceae exhibit a higher level of divergence than the exon regions. Gene site selection analysis suggested that 11 coding protein genes (*accD*, *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpoCl*, *rpoC2*, *rps16*, *ycfl*, and *ycf2*) have some sites under protein sequence evolution. Phylogenetic analysis showed that *G. huashanensis* is most closely related to the congeneric species *G. oldhamiana*. These results are very useful for studying phylogenetic evolution and species divergence in the family Caryophyllaceae.

ARTICLE HISTORY
Received 19 January 2023
Accepted 25 May 2023

KEYWORDS
Chloroplast genome; evolutionary selection; *Gypsophila huashanensis*; phylogenetic relationship

1. Introduction

Gypsophila huashanensis Y. W. Tsui & D. Q. Lu (Caryophyllaceae) is an endemic herb species in central China (Figure 1) (Lu 1994), and it is currently only distributed in the Huashan and Qinling Mountains in Shaanxi, China. *G. huashanensis* grows on mountain slopes, valleys, roadside grasslands, and rock crevices at 600–2600 m above sea level. Previous studies of species in the genus *Gypsophila* mainly focused on their chemical constituents and pharmacological effects (Xie et al. 2015; Zhu et al. 2016), whereas few have investigated genomic evolution in this genus. Acquiring chloroplast genome data is conducive to identifying further species and phylogenetic studies (Mehmood et al. 2020a, 2020b, 2020c).

2. Materials

The fresh *G. huashanensis* leaf tissues used in this study were sampled from the Qinling Mountains in China (108°55′23.115756″N, 34°14′58.102116″E, altitude 394.7 m). A plant voucher specimen (GHLZH2020113523) was deposited in the Laboratory of Plant Evolution and Ecology, Northwestern University (Xi'an, China) (Contact: Zhonghu Li, lizhonghu@nwu.edu.cn).

3. Methods

Total genomic DNA was isolated from *G. huashanensis* using a modified version of the hexadecyltrimethylammonium bromide

Figure 1. Plant characteristics image of *Gypsophila huashanensis*. The flower characteristics of *G. huashanensis* is the corymbose cymes terminal or borne in distal leaf axils, in subcapitate clusters; petals pinkish white, oblong-oblanceolate, ca. 5 mm, apex retuse; filaments exserted, linear, flat, unequal, shorter than to longer than petals, base broad. The photograph was taken by the authors in the Qinling Mountains (108°55′23.115756″N, 34°14′58.102116″E, altitude 394.7 m).

CONTACT Zhong-Hu Li lizhonghu@nwu.edu.cn Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an, Shaanxi, 710069, China; Min-Feng Fang fangmf@nwu.edu.cn Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Xi'an, Shaanxi, 710069, China

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2023.2220436

method (Doyle and Doyle 1990). After DNA quality and quantity testing, a paired-end library with an insert size of 350 bp was constructed and sequenced using the Illumina NovaSeq 6000 platform. The NGSQC Toolkit_v.2.3.3 was used to filter the raw sequencing reads (Patel and Jain 2012). *De novo* assembly was performed with SPAdes software (Bankevich et al. 2012). The assembly accuracy and efficiency were further improved using the GetOrganelle program (Jin et al. 2020). The circular plastome was obtained by using Bandage (Wick et al. 2015) and Geneious v.9.0.2 (https://www.geneious.com/) with *G. oldhamiana* (NC_058757) as the reference. The complete chloroplast genome of *G. huashanensis* was automatically annotated by PGA (Qu et al. 2019), and adjusted and confirmed in Geneious. Finally, a chloroplast genome map was drawn for *G. huashanensis* using CPGView (Figure 2) (Liu et al. 2023).

To infer the phylogenetic position of *G. huashanensis* in the family Caryophyllaceae, the complete chloroplast genome sequences of 21 plant species (including eight *Colobanthus* species, five *Pseudostellaria* species, three *Silene* species, and two outgroups belonging to Phytolaccaceae) were used to reconstruct their evolutionary relationships. First, the data matrices were aligned using the MAFFT v7 program (Katoh and Standley 2013). Second, maximum-likelihood (ML) and maximum parsimony (MP) phylogenetic trees were generated based on a concatenated data matrix of 21 complete chloroplast sequences. The ML tree was generated with the RAxML v8 program (Stamatakis 2014) under the GTR + G evolutionary model with 1000 bootstrap replicates. The MP tree was produced using the PAUP v.4 program (Swofford 2004) with 1000 bootstrap replicates.

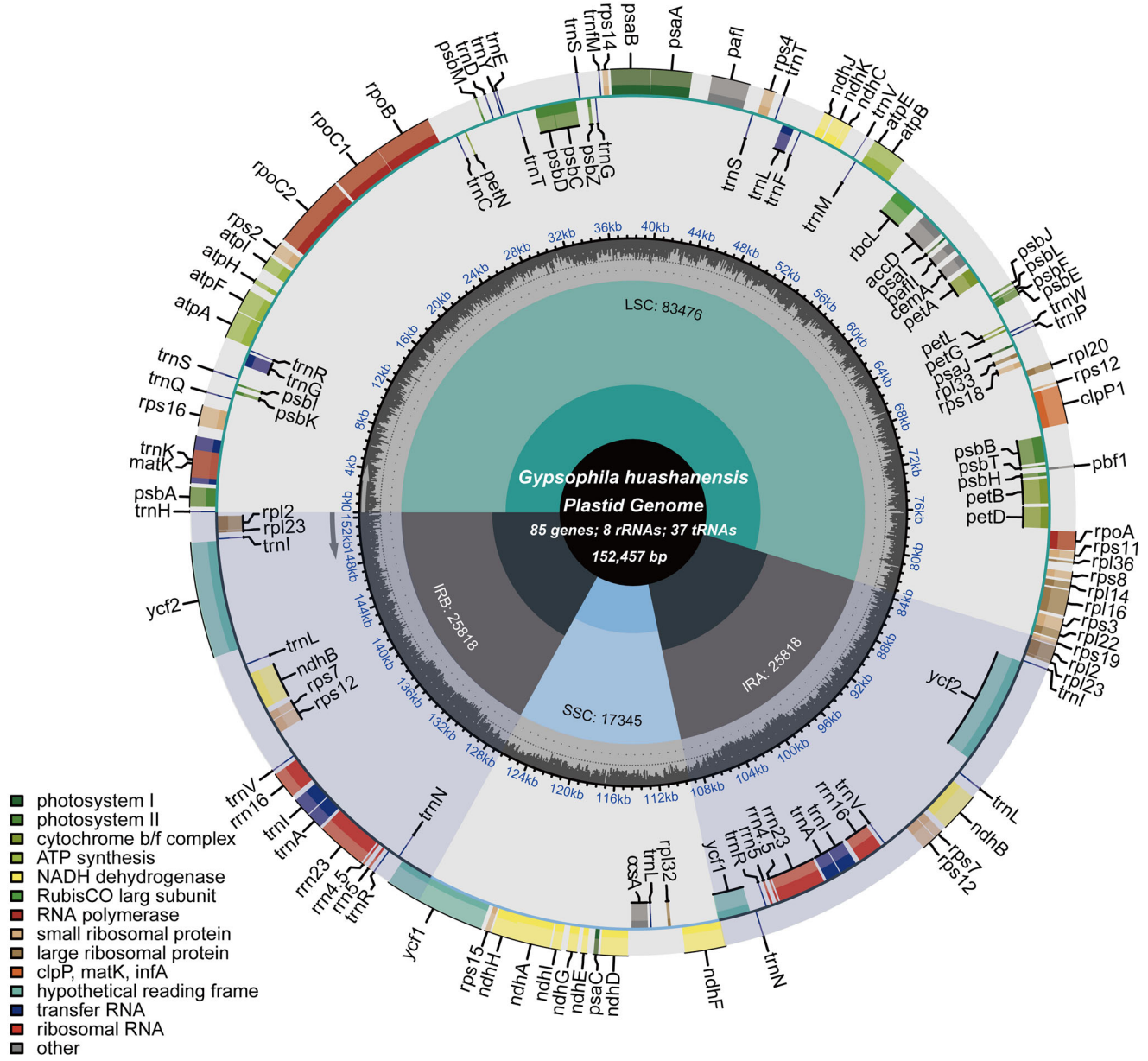

Figure 2. Circular map of the complete chloroplast genome of *Gypsophila huashanensis*. The center of the figure provides the specific information (genome length, GC content, and number of genes) of the *G. huashanensis* complete chloroplast genome sequence. From the center to the outside, the first track uses different colors to show the large single-copy (LSC) region (deep blue), small single-copy (SSC) region (light blue), and two inverted repeat (IRa and IRb) regions (gray). The GC content throughout the genome is plotted in the second track. Genes are indicated in the outermost track and color coded according to their functional classifications. The directions of transcription for the inner and outer genes are clockwise and anticlockwise, respectively. Different colors represent different gene types, the detailed gene types are listed in the captions.

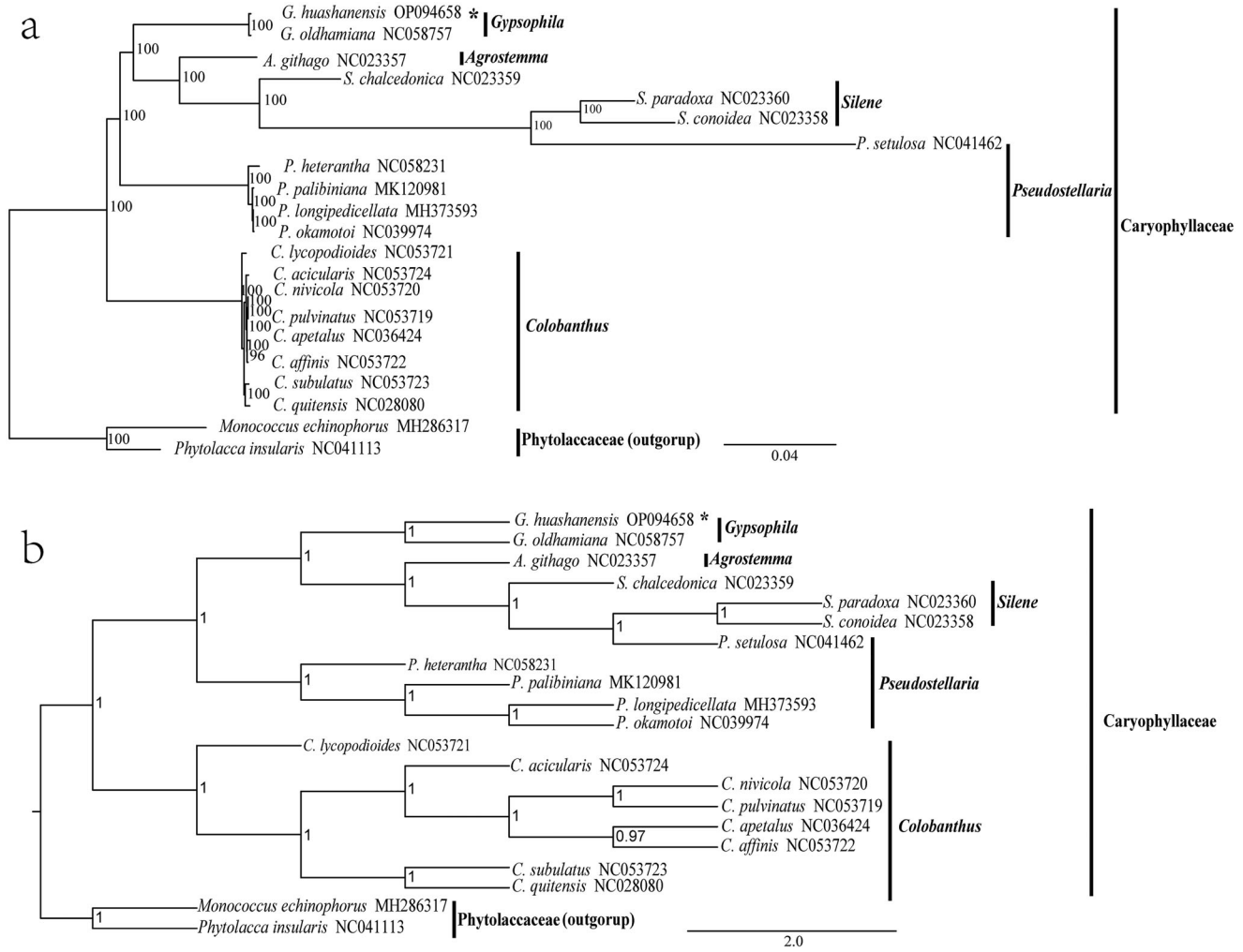

Figure 3. Phylogenetic relationships among *Gypsophila huashanensis* inferred from (a) maximum-likelihood (ML) method, and (b) maximum parsimony method based on concatenated complete chloroplast genome sequence of 21 species with two outgroups (Phytolaccaceae). *Newly sequenced plastid genome of *Gypsophila huashanensis*. The number on the branch represents bootstrap support. GenBank accession numbers of the following sequences were used: *G. oldhamiana* NC058757 (Jeong et al. 2021[a]), *A. githago* NC023357 (Sloan et al. 2014), *S. chalcedonica* NC023359 (Sloan et al. 2014), *S. paradoxa* NC023360 (Sloan et al. 2014), *S. conoidea* NC023358 (Sloan et al. 2014), *P. setulosa* NC041462 (Kim and Park 2019[a]), *P. heterantha* NC058231 (Kim et al. 2021[a]), *P. palibiniana* MK120981 (Kim et al. 2021[a]), *P. longipedicellata* MH373593 (Kim et al. 2021[a]), *P. okamotoi* NC039974 (Kim et al. 2019[a]), *C. lycopodioides* NC053721 (Androsiuk et al. 2020), *C. acicularis* NC053724 (Androsiuk et al. 2020), *C. nivicola* NC053720 (Androsiuk et al. 2020), *C. pulvinatus* NC053719 (Androsiuk et al. 2020), *C. apetalus* NC036424 (Androsiuk et al. 2017[a]), *C. affinis* NC053722 (Androsiuk et al. 2020), *C. subulatus* NC053723 (Androsiuk et al. 2020), *C. quitensis* NC028080 (Lee et al. 2015[a]), *Monococcus echinophorus* MH286317 (Yao et al. 2019), *Phytolacca insularis* NC041113 (Yang et al. 2019). [a]Direct submission to NCBI, unpublished.

In order to detect evolutionary variation in the chloroplast genomes in Caryophyllaceae, sequence alignment was conducted for the complete chloroplast genomes in Caryophyllaceae by using the mVISTA program (Frazer et al. 2004). We also performed gene site selection analysis for the protein-coding genes in Caryophyllaceae plastid genomes using the PAML 4.7 program (Yang et al. 2005). In order to choose the most reliable model, we have conducted likelihood ratio tests for evolutionary selection analysis.

4. Results and discussion

The complete chloroplast genome sequence of *G. huashanensis* (GenBank accession: OP094658) is 152,457 bp in length, with a GC content of 36.5% and it has four regions comprising two inverted repeat regions (IRs, 25,818 bp) separated by a large single-copy region (LSC, 83,476 bp) region and a small single-copy region (SSC, 17,345 bp). The read coverage depth map is shown in Supplementary Figure S1. The chloroplast genome contains 130 genes comprising 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. The GC contents of the chloroplast genome, LSC region, SSC region, and IR region are 34.1, 30.0, and 42.5%, respectively. Fourteen genes contain one intron (*rps16*, *atpF*, *rpoCl*, *petB*, *petD*, *rpl16*, *ndhB*, *ndhA*, *trnK-UUU*, *trnG-UCC*, *trnL-UAA*, *trnI-GAU*, *trnA-UGC*, and *trnI-GAU*), and three genes contain two introns (*rps12*, *pafI*, and *clpPl*). In addition, we mapped the structures of genes that are difficult to annotate in the chloroplast genome of *G. huashanensis* (Supplementary Figure S2).

Figure 3 shows that the ML and MP phylogenetic trees have similar topological structures. Phylogenetic analysis indicated that all of the Caryophyllaceae species considered in the present study clustered into a monophyletic evolutionary clade with high bootstrap support. *Gypsophila oldhamiana* was recovered as a sister taxon of *G. huashanensis*. The results obtained in this study extend our understanding of chloroplast genome evolution in the genus *Gypsophila*.

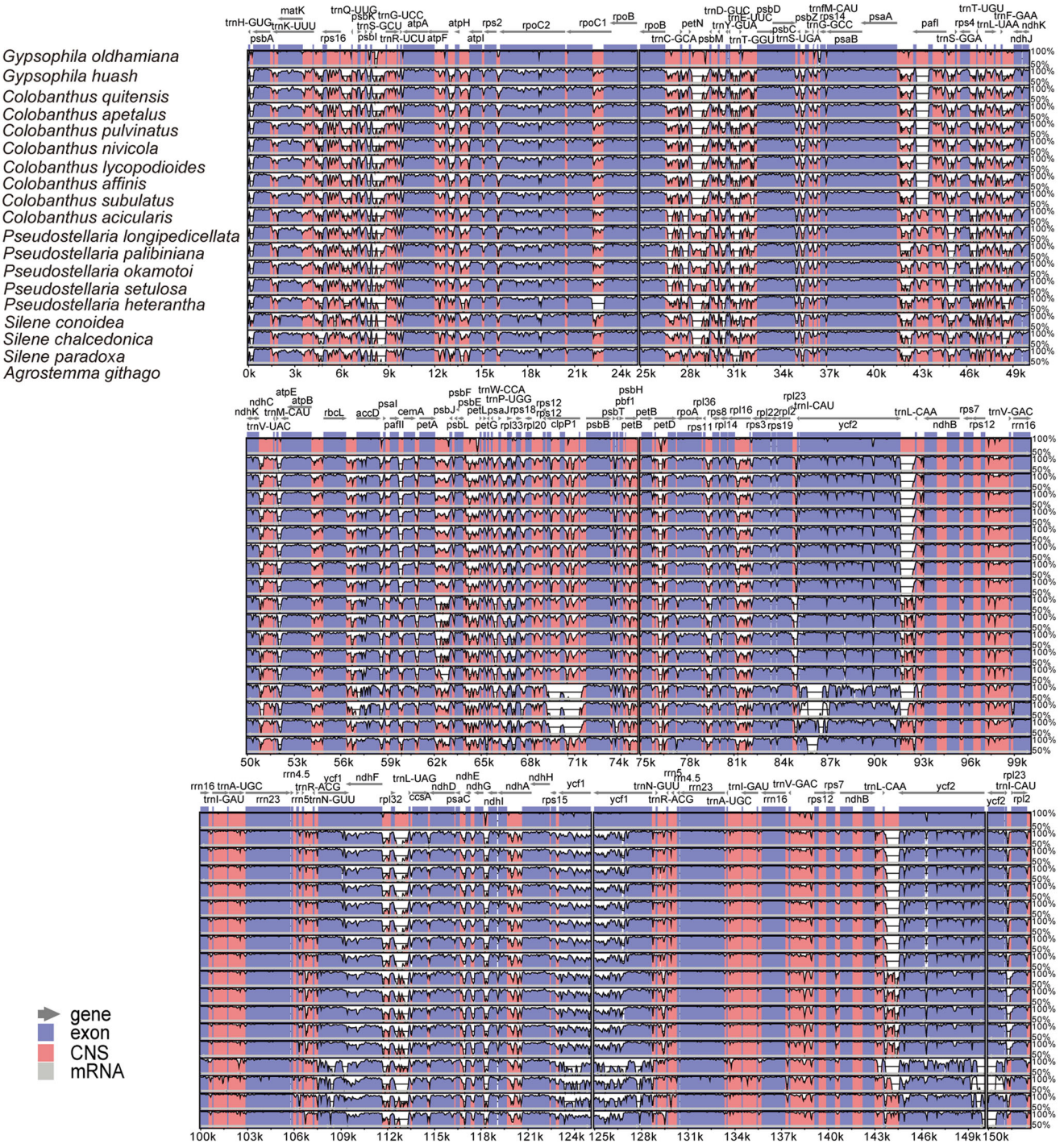

Figure 4. Sequence alignment of chloroplast genomes from 19 Caryophyllaceae species. Chloroplast genome sequences were aligned and compared with mVISTA software. The X-axis and Y-axis indicate the coordinates within the chloroplast genome and percentage identity (ranging from 50 to 100%), respectively. The grey arrows indicate the gene directions in the chloroplast genomes. Purple and pink bars represent exons and conserved non-coding sequences in chloroplast genomes, respectively.

Sequence evolution analysis showed that the non-coding regions of the chloroplast genomes of Caryophyllaceae species exhibit higher levels of genetic divergence than the exon regions (Figure 4). This result is consistent with the evolutionary characteristics of most angiosperm chloroplast genomes (Khakhlova and Bock 2006). We also detected 11 coding genes with some sites under positive selection ($p < 0.001$, Supplementary Table S1) comprising *accD*, *atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpoC1*, *rpoC2*, *rps16*, *ycf1*, and *ycf2*. In particular, the *accD*, *ndhA*, *petD*, *rpoC2*, *rps16*, *ycf1*, and *ycf2* genes were found to harbor multiple sites under evolutionary selection. The *accD* gene encoding acetyl-CoA carboxylase subunit is necessary for plant leaf development and it has important impacts on the leaf life and seed yield (Madoka et al. 2002; Kode et al. 2005). In addition, the *ndhA* gene encodes the NADH dehydrogenase subunit, which is involved in the electron transport chain and plant chlororespiration. The *petD* gene encodes cytochrome b6/f subunit IV, which plays important roles in linear and cyclic electron transport functions (Xiao et al. 2012). Moreover, the *ycf1* and *ycf2* genes are the largest genes in plastid genomes, and they encode part of the

chloroplast inner envelope membrane protein translocon (Kikuchi et al. 2013). These genes might have played important roles in environmental adaptation by *G. huashanensis*.

5. Conclusions

The complete chloroplast genome sequence of *G. huashanensis* was assembled and annotated in the present study. *G. oldhamiana* was found to be most closely related to *G. huashanensis*. Some genes under positive selection were identified in the chloroplast genome, and they might have played key roles in environmental adaptation by *G. huashanensis*. These results provide the basis for further studies of molecular evolution in Caryophyllaceae plants.

Ethical approval

Gypsophila huashanensis was not listed as a protected herb plant in China nor a threatened plant species on the IUCN Red List. Therefore, no specific permissions were needed for the sampling collections of *G. huashanensis* for scientific research purpose according to the regulations of the People's Republic of China on the protection and management of wild plants. During the field collecting process, we followed the local collecting guideline to ensure no substantial harm to the collecting wild plant individual.

Author contributions

Conception and design: Li ZH and Fang MF; software, analysis and interpretation of the data: Guan TX, Lu ZP, Liu ML, and Xun LL; the drafting of the paper, revising it critically for intellectual content: Guan TX, Lu ZP, and Liu ML; the final approval of the version to be published: Guan TX, Fang MF, and Li ZH. All authors agree to be accountable for all aspects of the work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was financially supported by the Key Program of Research and Development of Shaanxi Province [2022ZDLSF06-02].

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov/nuccore/OP094658/ under the accession no. OP094658. The associated BioProject, SRA, and Bio-Sample numbers are: PRJNA895000, SRR22100483, and SAMN31487758, respectively.

References

Androsiuk P, Jastrzębski JP, Paukszto Ł, Makowczenko K, Okorski A, Pszczółkowska A, Chwedorzewska KJ, Górecki R, Giełwanowska I. 2020. Evolutionary dynamics of the chloroplast genome sequences of six *Colobanthus* species. Sci Rep. 10(1):11522.

Bankevich A, Nurk S, Antipov D, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to singlecell sequencing. J Comput Biol. 19(5):455–77.

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from plant tissue. Focus. 12:13–15.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32(Web Server issue):W273–W279.

Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21(1):241.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46(1):85–94.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. 2013. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science. 339(6119):571–574.

Kode V, Mudd EA, Iamtham S, Day A. 2005. The tobacco plastid *accD* gene is essential and is required for leaf development. Plant J. 44(2):237–244.

Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Res. 23(3):694–704.

Lu DQ. 1994. The classification and distribution of *Gypsophila* (Caryophyllaceae) in China. Plant Res. 4:329–337.

Madoka Y, Tomizawa K-I, Mizoi J, Nishida I, Nagano Y, Sasaki Y. 2002. Chloroplast transformation with modified *accD* operon increases acetyl-coa carboxylase and causes extension of leaf longevity and increase in seed yield in tobacco. Plant Cell Physiol. 43(12):1518–1525.

Mehmood F, Shahzadi I, Ahmed I, Waheed MT, Mirza, B, Abdullah. 2020a. Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae. Genomics. 112(2):1522–1530.

Mehmood F, Ubaid Z, Bao Y, Poczai P, Mirza, B, Abdullah. 2020b. Comparative plastomics of Ashwagandha (*Withania*, Solanaceae) and identification of mutational hotspots for barcoding medicinal plants. Plants. 9(6):752.

Mehmood F, Ubaid Z, Shahzadi I, Ahmed I, Waheed MT, Poczai P, Mirza, B, Abdullah. 2020c. Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*). PeerJ. 8:e9552.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLOS One. 7(2):e30619.

Qu XJ, Moore MJ, Li DZ, Yi T S. 2019. P GA: a software package for rapid, accurate, and flexible batch annotation of plastomes. Plant Methods. 15:50–61.

Sloan DB, Triant DA, Forrester NJ, Bergner LM, Wu M, Taylor DR. 2014. A recurring syndrome of accelerated plastid genome evolution in the angiosperm tribe *Sileneae* (Caryophyllaceae). Mol Phylogenet Evol. 72:82–89.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and postanalysis of large phylogenies. Bioinformatics. 30(9):1312–1313.

Swofford DL. 2004. Paup 4.0 for Macintosh: phylogenetic analysis using parsimony (software and user's book for Macintosh). Sunderland (MA): Sinauer Associates, Incorporated.

Wick RR, Schultz MB, Zobel J, et al. 2015. Bandage: Interactive visualisation of de novo genome assemblies. Bioinformatics. 31(20):3350–3352.

Xiao J, Li J, Ouyang M, Yun T, He B, Ji D, Ma J, Chi W, Lu C, Zhang L, et al. 2012. DAC is involved in the accumulation of the cytochrome b6/f complex in *Arabidojpsis*. Plant Physiol. 160(4):1911–1922.

Xie LX, Sun DF, Wang HY, et al. 2015. Research progress on chemical constituents in plants of *Gypsophila* L. and their pharmacological activities. Chinese Tradit Herbal Drugs. 46(2):280–292.

Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol. 22(4):1107–1118.

Yang JY, Lee W, Pak J-H, Kim S-C. 2019. Complete chloroplast genome of Ulleung Island endemic pokeweed, *Phytolacca insularis* (Phytolaccaceae), in Korea. Mitochondrial DNA B Resour. 4(1):8–9.

Yao G, Jin J-J, Li H-T, Yang J-B, Mandala VS, Croley M, Mostow R, Douglas NA, Chase MW, Christenhusz MJ, et al. 2019. Plastid phylogenomic insights into the evolution of Caryophyllales. Mol Phylogenet Evol. 134:74–86.

Zhu BJ, Chen XR, Lu XZ. 2016. The analysis of nutritional components and amino acids in *Gypsophila oldhamiana* Miq. Hubei Agricul Sci. 55(15):3985–3987.