


The complete chloroplast genome of *Alhagi sparsifolia* Shap. (Leguminosae)

Xi Jiang^{a*}, Wen-Juan Huang^{b*}, Ya-Rui Li^{c*}, Jie Si^d, Jing-Dong Xu^c, Er-Dai Qin^c, Tian-Ge Yang^c, Hong Liu^c, Zhi-Hua Wu^{c*}  and Pei-Pei Jiao^{b,e*}

^aCollege of Plant Science, Tarim University, Alar, China; ^bXinjiang Production and Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, College of Life Science, Tarim University, Alar, China; ^cHubei Provincial Key Laboratory for Protection and Application of Special Plant Germplasm in Wuling Area of China, College of Life Sciences, South-Central University for Nationalities, Wuhan, China; ^dForeign Exchange and Cooperation Department, Tarim University, Alar, China; ^eCollege of Life Science and Technology, Huazhong Agricultural University, Wuhan, China

ABSTRACT

Alhagi sparsifolia Shap. is a perennial herbaceous plant belonging to the genus *Alhagi*, Leguminosae. This species is of high nutritional, medicinal and ecological values. The complete chloroplast genome was 128,418 bp and lost an IR (inverted repeat) region. Further annotation revealed the chloroplast genome contains 108 genes, including 75 protein coding genes, 29 tRNA genes, and 4 rRNA genes. A total of 103 simple sequence repeats (SSRs) were identified in the chloroplast genome. This chloroplast genome resource will be useful for study on the evolution and genetic diversity of *A. sparsifolia* in the future.

ARTICLE HISTORY

Received 31 January 2021
Accepted 12 June 2021

KEYWORDS

Alhagi sparsifolia;
Leguminosae; chloroplast;
genome; evolution

Alhagi sparsifolia Shap., is a perennial herbaceous plant belonging to the genus *Alhagi*, Leguminosae. It is widely distributed in arid desert areas of China, such as Inner Mongolia, Gansu, Qinghai and Xinjiang province. It provides plentiful crude protein, crude fiber, and unsaturated fatty acids for livestock as an ideal forage. A series of pharmacologically active secondary metabolites, such as flavonoids, alkaloids, steroids, pseudalhagin A, phospholipids, and polysaccharides have been extracted from *Alhagi* species (Muhammad et al. 2015), and they show broadly biological activities in antioxidant, cardiovascular, anti-ulcer, hepatoprotective, antispasmodic, antidiarrheal, antinociceptive, antipyretic, anti-inflammatory, anti-rheumatic, antibacterial, and antifungal (Muhammad et al. 2015; Zhou et al. 2017). Besides, *A. sparsifolia* also plays an important role in stabilizing and improving the fragile ecological environment in desert areas as a typical xerophyte. Therefore, *A. sparsifolia* has a series of nutritional, medicinal and ecological values.

In this study, to obtain a new insight into the phylogeny of *A. sparsifolia*, we sequenced, assembled, and annotated the accurate chloroplast genome of *A. sparsifolia*. The materials of *A. sparsifolia* in this study was collected from Wushi City, Aksu Prefecture, Xinjiang province of China (79°1'0.12"E, 41°7'1.8"N, 1580 m above the sea level). The complete genomic DNA was extracted using the modified CTAB method (Doyle and Doyle 1987) and sequenced using the Illumina NovaSeq platform at Majorbio Company (Shanghai, China).

The raw reads were generated and low-quality sequences were filtered out. The trimmed reads were assembled using GetOrganelle (Jin et al. 2020). The assembled genome was annotated using CPGAVAS2 (Shi et al. 2019) and PGA (Qu et al. 2019). The complete chloroplast genome was 128,418 bp (MW349013) and lost an IR region, the average GC content was 34.05%. Previous studies have examined the phylogenetic distribution of different plastid genome rearrangements among legumes, and found the loss of one copy of the IR from the tribes Carmichaelieae, Cicereae, Hedysareae, Trifolieae, Fabaeae, Galegeae, and three genera of Millettieae (Palmer and Thompson 1982; Lavin et al. 1990; Liston 1995; Jansen et al. 2008). The chloroplast genomes encoded 108 functional genes, including 75 protein coding genes, 29 tRNA genes, and 4 rRNA genes. A total of 103 SSR markers ranging from mononucleotide to hexa-nucleotide repeat motif were identified in *A. sparsifolia* chloroplast genome.

To explore the phylogenetic relationship of *A. sparsifolia* within Leguminosae, the complete chloroplast genomes of 23 species from Leguminosae were obtained from the GenBank database, with the *Polygala fallax* Hemsl. and *Polygala tenuifolia* Willd. as the outgroups, and the phylogenetic trees were built from the whole protein-coding gene matrix by maximum-likelihood (ML) and Bayesian inference (BI) (Figure 1). The ML tree was generated using IQ-TREE (Nguyen et al. 2015) based on the best model of TVM + F + I + G4 and 1000 bootstrap replicates, and BI

CONTACT Pei-Pei Jiao  zhwu@scuec.edu.cn, jiaopeipei2000@126.com  Xinjiang Production and Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, College of Life Science, Tarim University, Alar, China; College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China

*These authors have contributed equally to this work.

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

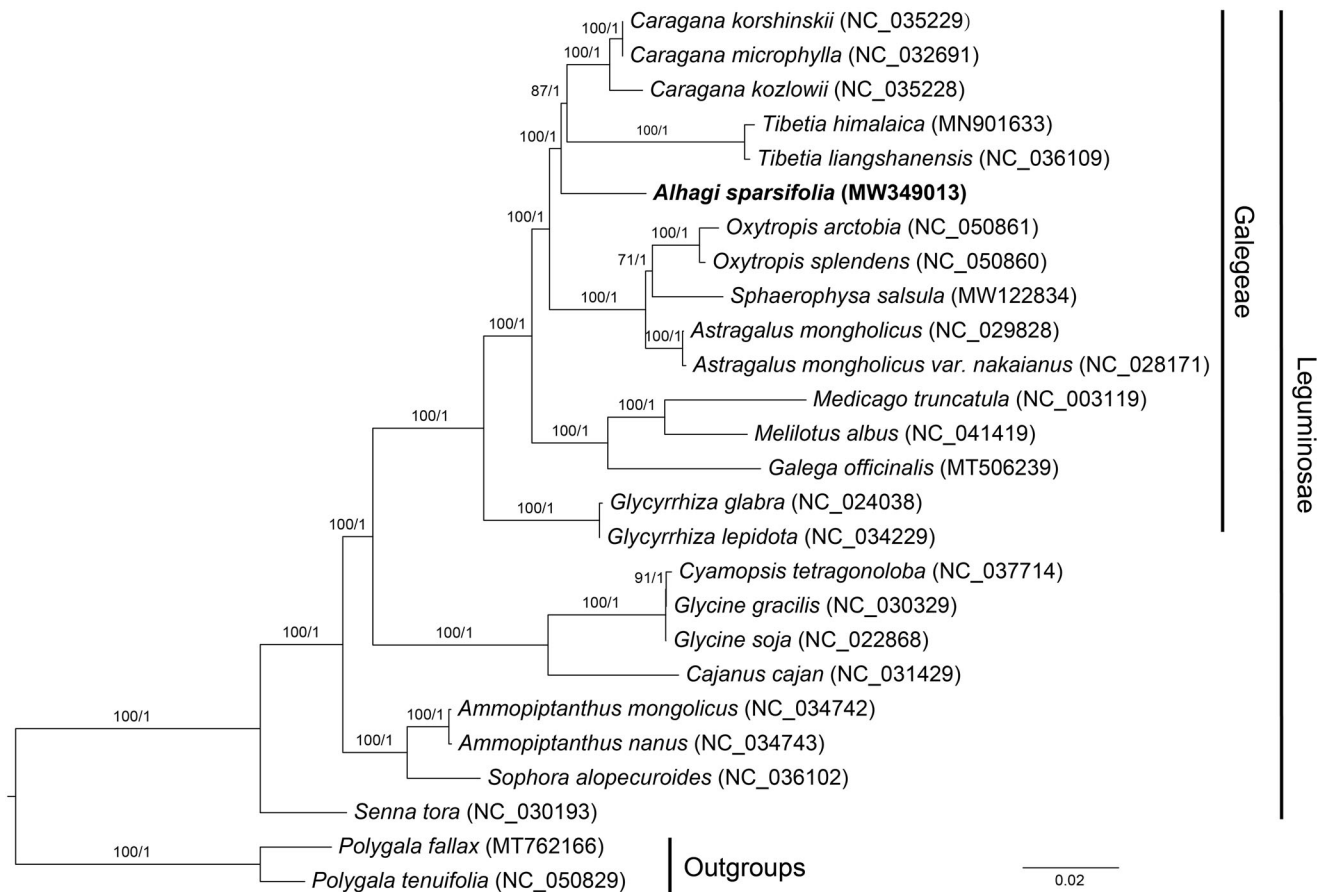


Figure 1. Phylogenetic tree reconstructed by maximum-likelihood (ML) and Bayesian inference (BI) analysis based on the whole chloroplast protein-coding genes of these 26 species. The numbers on each branch represent the bootstrap value for ML method and posterior probability for BI method, respectively.

analysis was performed in MrBayes-3.2.7 (Ronquist et al. 2012). The results show the analyzed *A. sparsifolia* was similar to other Galegeae species, which all lack an IR region. The *A. sparsifolia* was closer to the species of *Tibetia* and *Caragana*. This information will be useful for study on the evolution and genetic diversity of *A. sparsifolia* in the future.

Disclosure statement

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

Funding

This work was financially supported by the Joint Foundation of Huazhong Agricultural University & Tarim University research cooperation [TDHNLH201701], The Innovative Team Building Plan for Key Areas of Xinjiang Production and Construction Corps [2018CB003], Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin [BRZD2003].

ORCID

Zhi-Hua Wu  <http://orcid.org/0000-0002-2895-7256>

Data availability statement

The specimen and DNA sample were deposited at the herbarium of Tarim University (<https://www.taru.edu.cn/>), Wen-Juan Huang, hwjzky@

<https://www.ncbi.nlm.nih.gov/> under the voucher number TD-00558. 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/>] under the accession no. MW349013. The associated "BioProject", "SRA", and "Bio-Sample" numbers are PRJNA685338, SRR13255665, and SAMN17083245 respectively.

References

- Doyle JJ, Doyle JL. 1987. A Rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull.* 19:11–15.
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee SB, Daniell H. 2008. Complete plastid genome sequence of the chickpea (*Cicer arietinum*) and the phylogenetic distribution of rps12 and clpP intron losses among legumes (Leguminosae). *Mol Phylogenet Evol.* 48(3):1204–1217.
- 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.
- Lavin M, Doyle JJ, Palmer JD. 1990. Evolutionary significance of the loss of the chloroplast-DNA inverted repeat in the Leguminosae subfamily Papilionoidae. *Evolution.* 44(2):390–402.
- Liston A. 1995. Use of the polymerase chain reaction to survey for the loss of the inverted repeat in the legume chloroplast genome. In: Crisp M, Doyle JJ, editors. *Advances in legume systematics 7: phylogeny*. Kew: Royal Botanic Gardens; p. 31–40.
- Muhammad G, Hussain MA, Anwar F, Ashraf M, Gilani A. 2015. *Alhagi*: a plant genus rich in bioactives for pharmaceuticals. *Phytother Res.* 29(1):1–13.
- Nguyen LT, Schmidt HA, von Haeseler A, Quang MB. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.

- Palmer JD, Thompson WF. 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*. 29(2):537–550.
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15(1):50.
- Ronquist F, Teslenko M, Mark PVD, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542.
- Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 47(W1):W65–W73.
- Zhou D, Wei H, Jiang Z, Li X, Jiao K, Jia X, Hou Y, Li N. 2017. Natural potential neuroinflammatory inhibitors from *Alhagi sparsifolia* Shap. *Bioorg Med Chem Lett*. 27(4):973–978.