

## Complete plastid genome of *Lespedeza tricolor* (Fabaceae), an endemic shrub in Korea

Tae-Hee Kim , Sang-Chul Kim , Young-Ho Ha , Kyung Choi  and Hyuk-Jin Kim 

Division of Forest Biodiversity, Korea National Arboretum, Pocheon-si, Republic of Korea

### ABSTRACT

*Lespedeza tricolor* is a species found in the southern province of Korea, thought to be endemic to Korea. The complete plastid genome of this legume was sequenced in this study. DNA from *L. tricolor* was extracted, sequenced, and assembled into the complete plastid genome. We used 18 species of the family Fabaceae and 77 protein-coding genes to perform phylogenetic analysis. The plastid genome is 149,038 bp in length, with large (82,495 bp) and small (18,889 bp) single-copy regions, separated by a pair of inverted repeat regions (23,827 bp). It contains 83 protein-coding genes, eight rRNAs, 37 tRNAs, and two pseudogenes (*rpl22* and *infA*). Our phylogenetic analysis suggests that the genus *Lespedeza* is monophyletic and *L. tricolor* is closely related to *L. maritima* and *L. buergeri*. In this study, we identified the phylogenetic position of *L. tricolor* and provided the data that can be used in various ways in future studies.

### ARTICLE HISTORY

Received 9 June 2022  
Accepted 25 September 2022

### KEYWORDS

Plastid genome; *Lespedeza tricolor*; Korean endemic species; phylogenetic relationship

The genus *Lespedeza* Michx. (tribe Desmodieae; subfamily Papilionoideae; family Fabaceae), which comprises approximately 60 species, has the distribution of Asia and North America (Han et al. 2010). Due to their structure and characteristics, these species are used as forage and medicine, or for ornamental purposes (Cheng et al. 2004; Guan et al. 2013; Sharma and Rhyu 2015; Somaratne et al. 2019). In South Korea, 25 species have been recognized (Korea National Arboretum 2022). Among them, *Lespedeza tricolor* (Nakai) D. P. Jin, J. W. Park & B. H. Choi 2019 is an endemic species to Korea and inhabits the southern province (Jin et al. 2019). Many suggestions have been made on the taxonomic position of *L. tricolor*, such as, as a variety of *L. maximowiczii* [= var. *tricolor* Nakai], subspecies of *L. buergeri* [= subsp. *tricolor* (Nakai) Hatusima], or a synonym of *L. maximowiczii* (Hatusima 1967; Akiyama 1988; Jin et al. 2019). Jin et al. (2019) using microsatellite data analysis showed that although *L. tricolor* formed mixed lineages with *L. maximowiczii* and *L. buergeri*, it was a distinct species. To clarify the phylogenetic relationships of *L. tricolor* and related species, this study aims to determine the complete plastid genome sequence of *L. tricolor* and conduct phylogenetic analysis using 77 plastid protein-coding gene datasets.

Fresh leaves of *L. tricolor* were collected from Wando-gun, Jeollanam-do province, South Korea (34°20'25.0"N 126°41'37.6"E) and dried directly with silica gel at room temperature until DNA extraction. The voucher specimen was deposited in the herbarium of the Korea National Arboretum (KH; [http://www.nature.go.kr/kbi/plant/smpl/KBI\\_2001\\_030100.do](http://www.nature.go.kr/kbi/plant/smpl/KBI_2001_030100.do), Hee

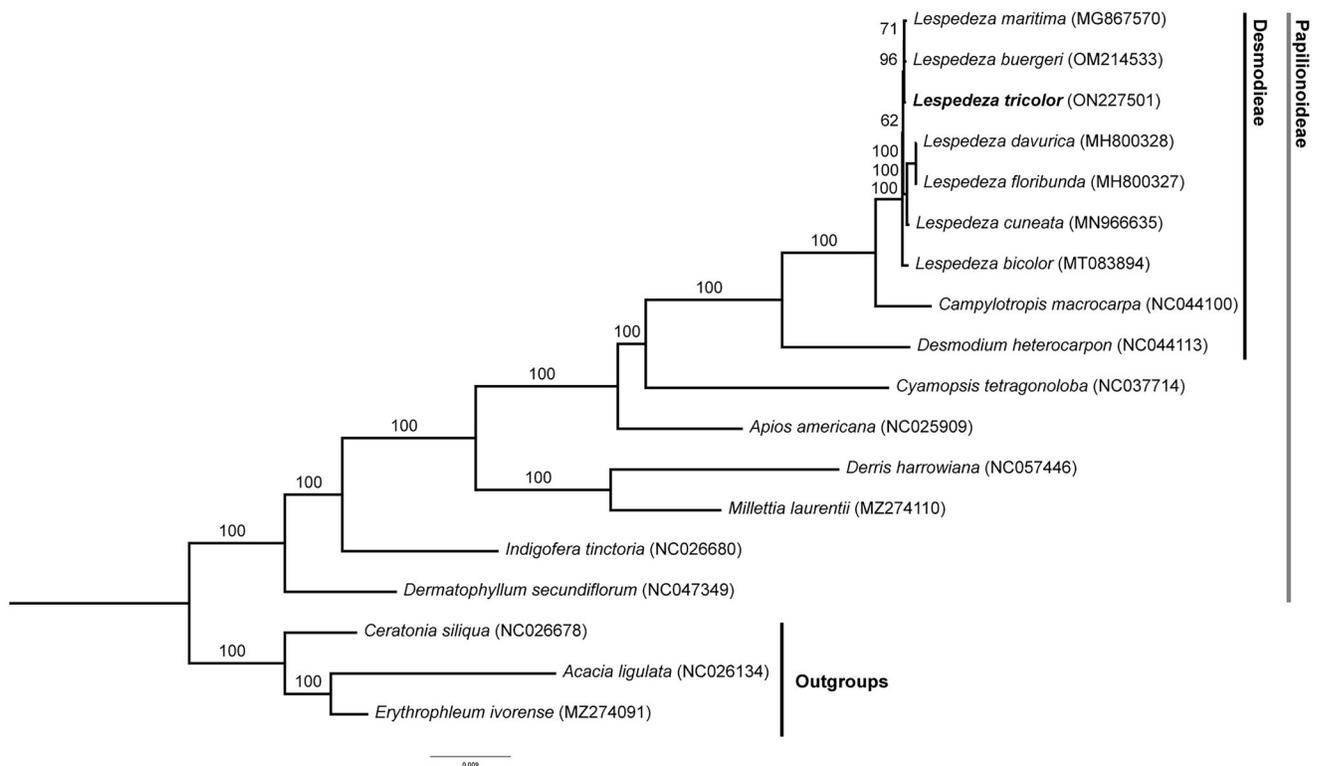
Young Gil, E-mail: [warmishe@korea.kr](mailto:warmishe@korea.kr), voucher number: ESK21-503). Total genomic DNA was extracted using DNeasy Plant Mini Kit (Qiagen Inc., Valencia, CA). Next-generation sequencing was conducted using MiSeq sequencing system (Illumina, Seoul, South Korea) and a total of 10,049,414 reads were obtained. The GetOrganelle software was used to select the plastid-like reads, which were assembled into the complete plastid genome by Geneious Prime program and the GetOrganelle toolkit (Kearse et al. 2012; Jin et al. 2020). Gene content and order were annotated using the Geseq tool and Geneious Prime (Kearse et al. 2012; Tillich et al. 2017).

To identify the phylogenetic relationships of subfamily Papilionoideae, a total of 18 species were used, of which three species (*Acacia ligulata*, GenBank accession = NC\_026134; *Erythrophleum ivorense*, GenBank accession = MZ274091; *Ceratonia siliqua*, GenBank accession = NC\_026678) were designated as outgroups. For phylogenetic analysis, 77 protein-coding genes were aligned using the MAFFT program and maximum-likelihood (ML) analysis in the software PhyloSuite (Kato and Standley 2013; Zhang et al. 2020). The best substitution model was TVM + F+R2 according to ModelFinder in PhyloSuite (Zhang et al. 2020). 1000 bootstrap (BS) replications were performed based on a dataset of protein-coding genes.

The complete plastid genome of *L. tricolor* (GenBank accession = ON227501) is 149,038 base pairs (bp) long, with a typical quadripartite structure which has components such as a large single-copy (LSC) of 82,495 bp, a small single-copy (SSC) of 18,889 bp, and two inverted repeats (IRs) of 23,827 bp. The total GC content was 35.0%, and the GC

**CONTACT** Hyuk-Jin Kim  [jins77@korea.kr](mailto:jins77@korea.kr)  Division of Forest Biodiversity, Korea National Arboretum, 415 Gwangneungsumogwon-ro, Soheul-eup, Pocheon, Gyeonggi-do 11186, Republic of Korea

© 2022 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.** Maximum-likelihood tree of 18 legumes inferred from combining 77 protein-coding genes. The numbers above each node are the bootstrap support values from 1000 replications.

contents of the LSC, SSC, and IR regions were 32.4%, 28.2%, and 42.2%, respectively. The plastid genome includes 128 genes (83 protein-coding genes, eight rRNAs, and 37 tRNAs) including 17 genes (six protein-coding genes, four rRNAs, and seven tRNAs) duplicated in the IR regions. Two genes (*rpl22* and *infA*) were missed in *L. tricolor* (Millen et al. 2001; Jansen et al. 2008; Magee et al. 2010). In addition, the tribe Desmodieae is known to have no introns within *rps12* and *rpl2* (Jansen et al. 2008). This study also confirmed the absence of introns for these genes in *L. tricolor*.

The ML trees showed that Papilionoideae were monophyletic with high BS support values (100) and *L. tricolor* was a sister group of *L. maritima* and *L. buergeri* (Figure 1). The complete plastid genome of *L. tricolor* will be useful to study the phylogeny, develop identification markers, and understand the evolutionary history for *Lespedeza* species.

### Ethical approval

The material involved in the paper does not involve ethical conflicts. Therefore, it requires no specific permissions or licenses.

### Author contributions

Hyuk-Jin and Kyung conceived the original structure of the review. Sang-Chul and Young-Ho have collected the data and conducted the experiments. Tae-Hee analyzed the data and drafted the manuscript. All authors have read and agreed to the submitted version of the manuscript.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This study was supported by the Scientific Research of the Korea National Arboretum under Grant number [KNA1-1-13, 14-1].

### ORCID

Tae-Hee Kim  <http://orcid.org/0000-0002-1727-1435>  
 Sang-Chul Kim  <http://orcid.org/0000-0003-0698-1500>  
 Young-Ho Ha  <http://orcid.org/0000-0002-2374-3611>  
 Kyung Choi  <http://orcid.org/0000-0003-0782-9324>  
 Hyuk-Jin Kim  <http://orcid.org/0000-0002-3177-2914>

### Data availability statement

The data that support the findings of this study are openly available in NCBI at <https://www.ncbi.nlm.nih.gov/> (reference number ON227501). The associated BioProject, SRA, and Bio-Sample numbers are PRJNA826218, SRR18762682, and SAMN27553414, respectively.

### References

- Akiyama S. 1988. A revision of the genus *Lespedeza* section *Macrolespedeza* (Leguminosae). Bull Univ Mus Univ Tokyo. 33:1–170.
- Cheng X, An S, Liu S, Li G. 2004. Micro-scale spatial heterogeneity and the loss of carbon, nitrogen and phosphorus in degraded grassland in Ordos Plateau, northwestern China. Plant Soil. 259(1/2):29–37.
- Guan XK, Zhang XH, Turner N, Xu BC, Li FM. 2013. Two perennial legumes (*Astragalus adsurgens* Pall. and *Lespedeza davurica* S.) adapted to semiarid environments are not as productive as lucerne

- (*Medicago sativa* L.), but use less water. *Grass Forage Sci.* 68(3): 469–478.
- Han JE, Chung K-H, Nemoto T, Choi B-H. 2010. Phylogenetic analysis of eastern Asian and eastern North American disjunct *Lespedeza* (Fabaceae) inferred from nuclear ribosomal ITS and plastid region sequences. *Bot J Linn Soc.* 164(3):221–235.
- Hatusima S. 1967. *Lespedeza*: sects. *Macrolespedeza* and *Heterolespedeza* from Japan, Korea and Formosa. *Mem Fac Agr Kagoshima Univ.* 6(1): 1–17.
- Jansen RK, Wojciechowski MF, Sanniyasi E, Lee S-B, 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 D-P, Park J-W, Choi B-H. 2019. Taxonomic study on infraspecific taxa of *Lespedeza maximowiczii* and hybrids with related species. *Korean J Plant Taxon.* 49(4):300–318.
- 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):1–31.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markovitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28(12): 1647–1649.
- Korea National Arboretum. 2022. Checklist of vascular plants in Korea (native plants). Pocheon: Korea National Arboretum.
- Magee AM, Aspinall S, Rice DW, Cusack BP, Sémon M, Perry AS, Stefanović S, Milbourne D, Barth S, Palmer JD, et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. *Genome Res.* 20(12):1700–1710.
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanagh TA, Hibberd JM, Gray JC, Morden CW, et al. 2001. Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell.* 13(3): 645–658.
- Sharma BR, Rhyu DY. 2015. *Lespedeza davurica* (Lax.) Schindl. extract protects against cytokine-induced  $\beta$ -cell damage and streptozotocin-induced diabetes. *Biomed Res Int.* 2015:169256.
- Somaratne Y, Guan D-L, Wang W-Q, Zhao L, Xu S-Q. 2019. The complete chloroplast genomes of two *Lespedeza* species: insights into codon usage bias, RNA editing sites, and phylogenetic relationships in Desmodieae (Fabaceae: Papilionoideae). *Plants.* 9(1):51.
- 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(W1):W6–W11.
- Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 20(1):348–355.