

The complete chloroplast genome sequence of a vulnerable legume species *Dalbergia tonkinensis* prain in Vietnam

Changyoung Lee^{a,b*}, Yong-In Kim^{c*}, Soo-Yong Kim^a, Tran The Bach^d, Sangmi Eum^a, Nguyen Trung Thanh^b and Sangho Choi^a

^aInternational Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea; ^bDepartment of Botany, VNU University of Science, Hanoi, Vietnam; ^cMultidisciplinary Genome Institute, Life Science Hall, Hallym University, Chuncheon, Republic of Korea; ^dInstitute of Ecology and Biological Resources (IEBR), Vietnam Academy of Science and Technology (VAST), Hanoi, Vietnam

ABSTRACT

Dalbergia tonkinensis is a critically vulnerable tree species that is distributed in Vietnam and Hainan Island of China. The complete chloroplast genome sequence of *D. tonkinensis* was characterized using Illumina pair-end sequencing. The cpDNA is 156,086 bp in length and contains a pair of 25,720 bp inverted repeats, one large single copy region of 85,761 bp, and one small single copy region of 18,885 bp. It contains 131 genes including 86 protein-coding genes, 36 tRNAs, eight rRNAs, and one pseudogene. The overall G + C content of the whole genome is 36.1%. Phylogenetic analysis based on 35 chloroplast genomes of Papilionoideae including *Cercis glabra* (as an outgroup) indicates that all the species of the Dalbergioids *sensu lato* formed a monophyletic clade and *D. tonkinensis* formed a sister relationship with the *D. hainanensis* and *D. odorifera* group.

ARTICLE HISTORY

Received 17 June 2019
Accepted 22 June 2019

KEYWORDS



Dalbergia tonkinensis;
vulnerable species;
chloroplast genome;
phylogenetic analysis

Dalbergia tonkinensis is a 5–13 m tall floral tree species belonging to Papilionoideae de Candolle (Fabaceae) and is distributed in Vietnam and Hainan Island of China (Nguyen et al. 2018). This plant was encountered the serious threat as the results of human activities and overexploitation for economic uses. It has been classified as a critically vulnerable plant listed in the Vietnam Red List (Dang and Nguyen 2007) and IUCN Red list (Ban 1998). In this study, the complete chloroplast genome of *D. tonkinensis* was reconstructed based on Illumina sequencing to provide the underlying information for genetic and conservation studies and explore possible phylogenetic relationship within Papilionoideae.

The plant material was sampled at Bac Huong Hoa nature reserve (Vietnam). The voucher specimen (KRIB0077649) was deposited in herbarium of KRIB. The total genomic DNA was sequenced using the Illumina Hiseq 2500 platform. The pair-end library was comprised 43,285,230 reads (length 2×101 bp). The Geneious *de novo* assembler was used to assemble the trimmed paired-read set. generated 565,298 contigs. The draft chloroplast genome of 156,086 bp achieved by identifying the repetitive inverted repeats sequences at 131,558 bp contig suspected of chloroplast genome with a

mean coverage of 287.8 X. The gene annotations were conducted by using Geneious v 11.1.5 (Biomatters Ltd., New Zealand) via comparison with chloroplast genomes of *D. odorifera* T. Chen (MF668133) and *D. hainanensis* Merr. et Chun (MF926268). The annotated chloroplast genome sequence was submitted to the GenBank (MK599470).

The complete chloroplast genome of *D. tonkinensis* has a total length of 156,086 bp, with a pair of inverted repeats of 25,720 bp that separate a large single copy region of 85,761 bp and a small single copy region of 18,885 bp. The chloroplast genome contained 131 genes including 86 protein-coding genes, 36 tRNA genes and 8 ribosomal RNA genes. Among these genes, 15 genes (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12*, *ycf68*, *trnA-UGC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*) have one intron and three genes (*clpP*, *rps12*, *ycf3*) have two introns. The *ycf1* gene was inferred to be pseudogenes. Most of the genes occurred as a single copy; however, six protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *ycf2*), seven tRNA genes (*trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*) and four rRNA genes (*rrn16*, *rrn23*, *rrn4.5*, *rrn5*) in the IR regions were duplicated. The overall G + C content of chloroplast genome is 36.1%, respectively.

CONTACT Sangho Choi  decoy0@kribb.re.kr  International Biological Material Research Center, Korea Research Institute of Bioscience and Biotechnology, 125, Gwahak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

*Changyoung Lee and Yong-In Kim have contributed equally to this work.

© 2019 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.

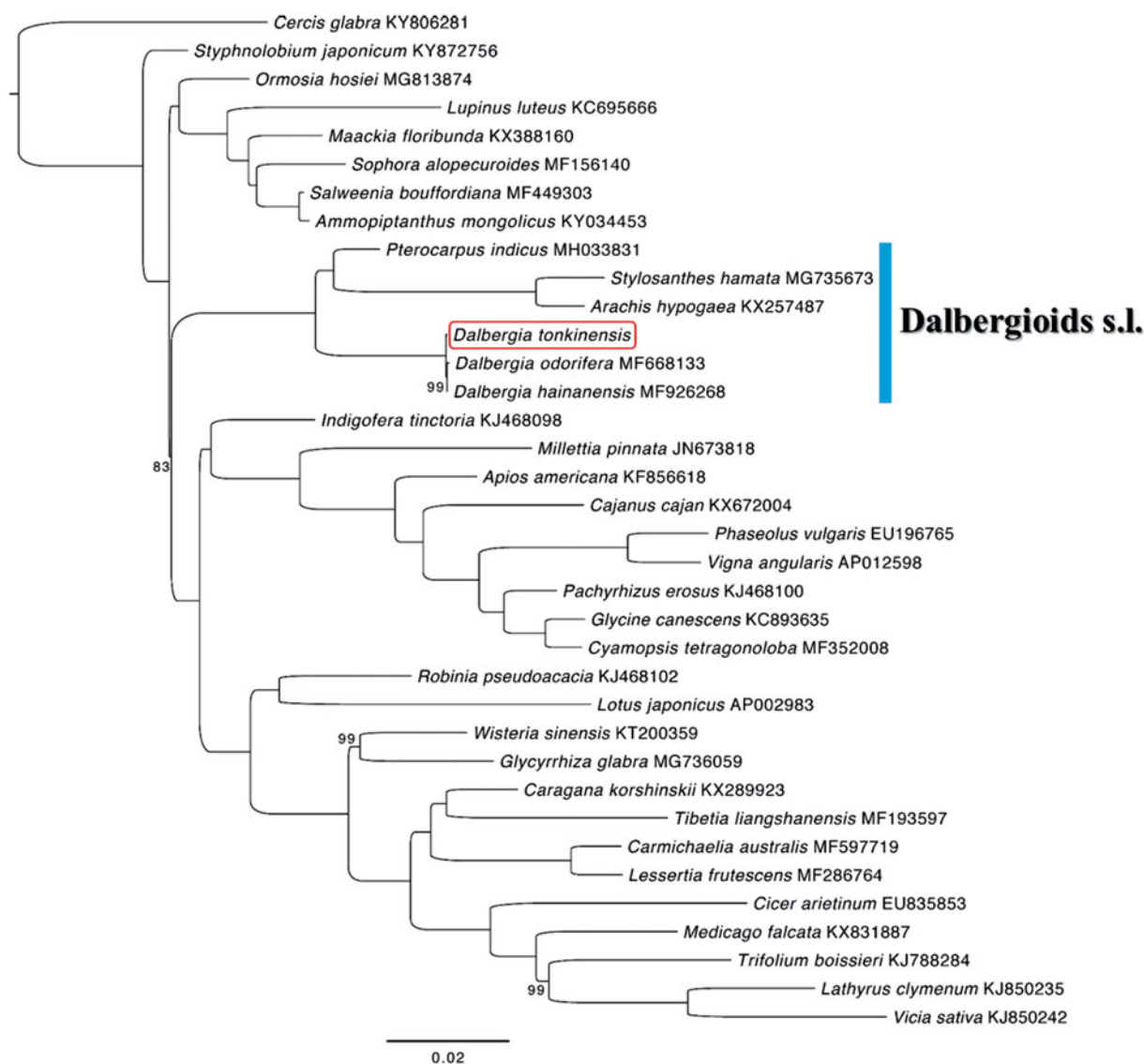


Figure 1. Maximum-likelihood phylogenetic tree based on 35 complete chloroplast genomes of Papilionoideae including *Cercis glabra*, as an outgroup. The number on each node indicates the bootstrap value less than 100. The position of *Dalbergia tonkinensis* is shown in a red blank box.

Phylogenetic analysis was performed using chloroplast coding sequences of *D. tonkinensis* and those of 35 related species of Papilionoideae including *Cercis glabra* Pamp, as an outgroup. These chloroplast genomes were aligned by using MAFFT (Kato and Standley 2013). Maximum-likelihood tree was generated using IQ-TREE (Nguyen et al. 2015) based on TVM + F + R3 model with ML + ultrafast bootstrap (Minh et al. 2013) 1000 replicates. As shown in the highly resolved ML phylogenetic tree (Figure 1), all the species of the Dalbergioids *sensu lato* formed a monophyletic clade and *D. tonkinensis* formed a sister relationship with the *D. hainanensis* and *D. odorifera* group.

Acknowledgement

The authors would like to thank Mr. Dao Quang Canh (Director) and Mr. Ha van Hoan (Vice Director) of BHHNR for collecting sample.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by Korea Research Institute of Bioscience and Biotechnology Research Initiative Program of the Republic of Korea.

References

- Ban NT. 1998. *Dalbergia tonkinensis*. The IUCN Red List of Threatened Species 1998: e.T32819A9732061.
- Dang NT, Nguyen TB. 2007. Vietnam Red List. Publishing House for Science and Technology Hanoi 412.
- Kato K, Standley DM. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.
- Minh BQ, Nguyen MAT, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30:1188–1195.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268–274.
- Nguyen VB, Wang S-L, Nhan NT, Nguyen TH, Nguyen NPD, Nghi DH, Cuong NM. 2018. New records of potent in-vitro antidiabetic properties of *Dalbergia tonkinensis* heartwood and the bioactivity-guided isolation of active compounds. *Molecules.* 23:1589.