

The complete chloroplast genome of *Keteleeria davidiana* var. *calcareo* (Pinaceae), an endangered species endemic to China

Jingjian Li^a, Niya Cheng^a and Yancai Shi^b

^aCollege of Pharmacy, Guilin Medical University, Guilin, China; ^bThe Chinese Academy of Sciences, Guangxi Institute of Botany, Guilin, China

ABSTRACT

Keteleeria davidiana var. *calcareo* is an endangered tree with considerable economic potential that used as timber wood for furniture and house construction. However, the natural population of *K. davidiana* var. *calcareo* is very fragmented, which is the cause for its low genetic diversity. In this study, we report the complete chloroplast genome of *K. davidiana* var. *calcareo* using Illumina sequencing. The chloroplast genome size is 117,670 bp in length, harboring a pair of very short inverted repeats (IRs) of 262 bp separated by a large single copy (LSC) sequence of 64,634 bp and a small single copy (SSC) sequence of 53,078 bp. The chloroplast genome *K. davidiana* var. *calcareo* contains 113 genes (74 protein genes, 35 tRNA genes, and 4 rRNA genes) and the overall GC content is 38.6%. The maximum likelihood phylogenetic analysis shows that *K. davidiana* var. *calcareo* is clustered with *K. davidiana* in genus *Keteleeria*. This complete chloroplast genome will help us to understand the evolution of *K. davidiana* var. *calcareo* and lays the foundations for future studies in this species conservation.

ARTICLE HISTORY

Received 9 November 2020
Accepted 24 January 2021

KEYWORDS

Keteleeria davidiana var. *calcareo*; chloroplast genome; phylogenetic analysis

Keteleeria davidiana var. *calcareo* is a rare tree species which grows mainly in southwestern China. This tree exhibits strong and heavy wood material characteristic, which make it suitable for house construction, furniture manufacture and ship-building (Jiang et al. 2008). At present, due to its vulnerable reproductive capacity couple with excessive logging by human, *K. davidiana* var. *calcareo* has become an endangered species and has been listed as important protected species in China (Xie et al. 2017). In order to protect this species effectively, researchers have paid a lot attention to formulate strategies for the cultivation and conservation of *K. davidiana* var. *calcareo* (Xie et al. 2017). However, there are still few studies regarding to the evolution and phylogeny of *K. davidiana* var. *calcareo* from the molecular level, which we believe to be an important yet underappreciated direction for this species conservation. A complete chloroplast genome sequence of *K. davidiana* var. *calcareo* will contribute to the progress of these works.

Chloroplast genome is often utilized for phylogenetic analysis and domestication studies of higher plants (Nie et al. 2020). The whole chloroplast genome sequences have also been demonstrated the potential to understand structure and functional evolution (Sabater 2018; Cheon et al., 2019; Zha et al. 2020). In the genus *Keteleeria*, the chloroplast genome of some species such as *K. davidiana* and *K. evelyniana* has been reported (Wu et al. 2009; Li et al. 2019), but the chloroplast genome of *K. davidiana* var. *calcareo* has not been reported. Here, we sequenced and analyzed the complete chloroplast genome sequence of *K. davidiana* var.

calcareo based on the Illumina sequencing data. This study aimed to characterize the complete chloroplast genome sequence of *K. davidiana* var. *calcareo* as a resource for future genetic studies.

Fresh young leaves from one individual of *K. davidiana* var. *calcareo* were collected from Guangxi Institute of Botany, The Chinese Academy of Sciences, Guilin, China (latitude: 25.0677; longitude: 110.3037). The total DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN, German). Voucher specimen of *K. davidiana* var. *calcareo* was deposited at the herbarium of Guangxi Institute of Botany (contact person: Yancai Shi, email: shiyancainan@163.com) under the voucher number IBK-SYC-202006). The whole-genomic DNA data was sequenced using the Illumina HiSeq2000 platform, which was then assembled using the program GetOrganelle (Jin et al. 2020) with *K. davidiana* (NC_011930) as reference. The assembled chloroplast genome was annotated by the combination of PGA (Qu et al. 2019) and GeSeq (Tillich et al. 2017). For necessary genes, we manually corrected their positions of start and stop codons and boundaries between exons and introns.

The complete chloroplast genome of *K. davidiana* var. *calcareo* is 117,670 bp in length with a typical quadripartite structure contains a pair of short inverted repeat regions (IRa and IRb) consisting of 262 bp each, a large single copy (LSC) region of 64,634 bp and a small single copy (SSC) sequence of 53,078 bp. The chloroplast genome contains a total of 113 genes, including 74 protein-coding genes, 35 tRNA genes and 4 rRNA genes. The gene content and gene order are

CONTACT Yancai Shi  shiyancainan@163.com  The Chinese Academy of Sciences, Guangxi Institute of Botany, Guilin, China

© 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.

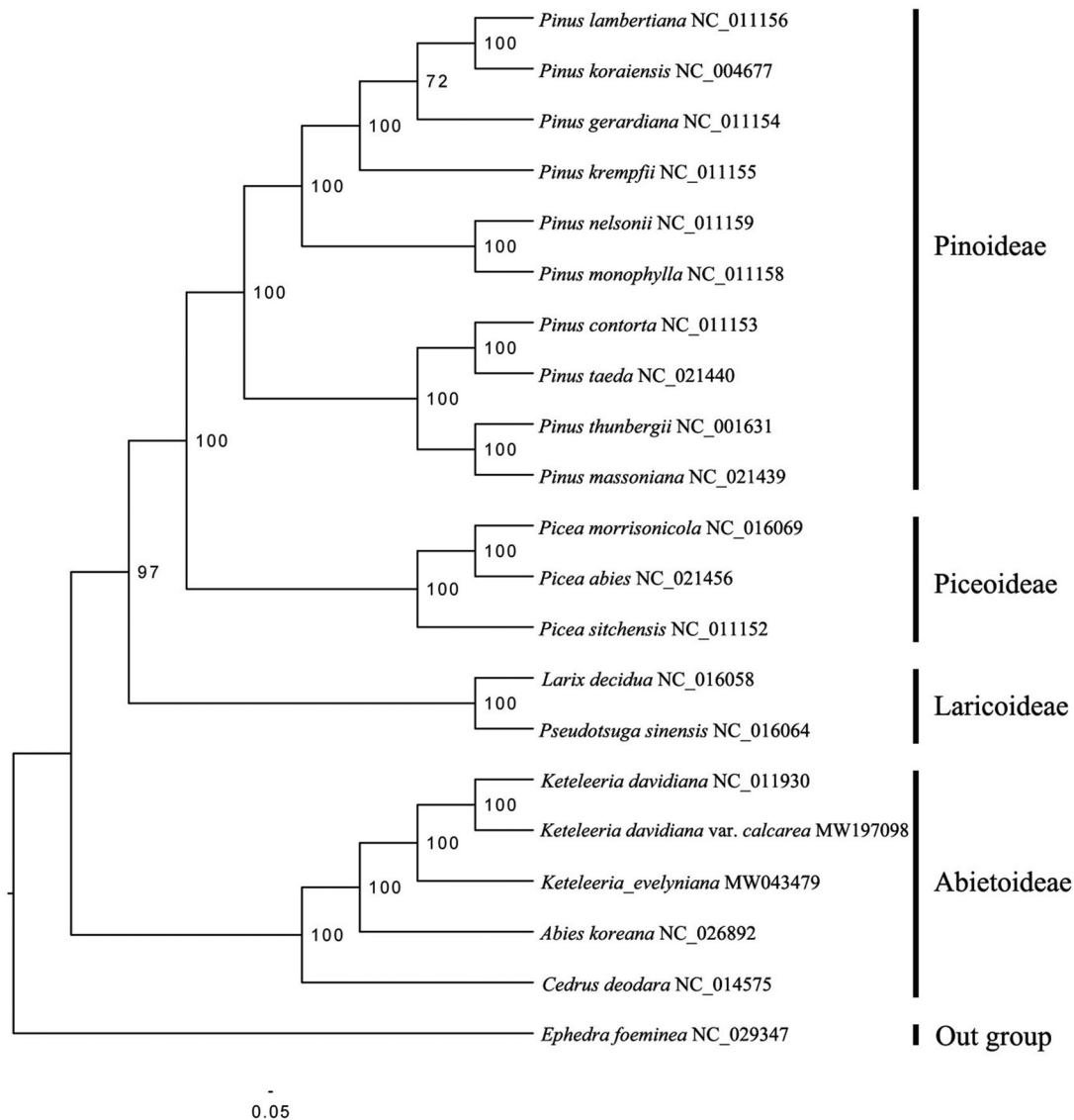


Figure 1. Maximum likelihood phylogenomic tree inferred from *K. davidiana* var. *calcareae* and other 19 species within Pinaceae and one out group species (*E. foeminea*) using complete chloroplast genomes. Bootstrap supports based on 1000 replicates are given at the node.

similar to the chloroplast genomes of other species in Pinaceae (Yi et al. 2016). Fifty-seven protein coding and 15 tRNA genes are located in the LSC region, while 17 protein-coding, 18 tRNA genes and 4 rRNA are located in the SSC region, respectively. Only one tRNA gene (trnI-CAU) is duplicated and located on the IR regions. All *ndh* genes have been lost in the genome of *K. davidiana* var. *calcareae* like other chloroplast genomes of family Pinaceae. The overall GC content of *K. davidiana* var. *calcareae* chloroplast genome is 38.6%, which is similar with those of other member of family Pinaceae and general angiosperms (Qian et al. 2013; Yi et al. 2016).

To confirm the phylogenetic position of *K. davidiana* var. *calcareae*, a phylogenomic analysis was performed based on 19 published species within Pinaceae and one out group (*Ephedra foeminea*, Ephedraceae). A total of seventy-two protein coding genes shared by all the species were extracted, and were aligned by using MUSCLE (Edgar 2004). Tree topologies for Maximum likelihood (ML) analysis using RAxML 8.0 software (Stamatakis 2014) based on the GTR+R+I model

were congruent with each other and all clades were strongly supported in those trees (–lnL of –192708.384031). The ML tree supported that *K. davidiana* var. *calcareae* and *K. davidiana* cluster as sister in genus *Keteleeria* (Figure 1), which is consistent with the previous studies in Pinaceae (Li et al. 2019; Yi et al. 2016). In conclusion, this published *K. davidiana* var. *calcareae* chloroplast genome will provide a solid foundation for phylogenetic and evolutionary studies in *Keteleeria* and is expected to help us formulate protection strategy for this species conservation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Natural Science Foundation of China [32060351] and Special Fund of Guangxi Talent Highlands.

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/>) under the accession no. MW197098. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA674712, SUB8472759, and SAMN16672654, respectively. Tree file of 21 species and genes for phylogenetic analysis were deposited at Figshare: <https://doi.org/10.6084/m9.figshare.13207604>.

References

- Cheon KS, Kim KA, Kwak M, Lee B, Yoo KO, Chiang TY. 2019. The complete chloroplast genome sequences of four *Viola* species (Violaceae) and comparative analyses with its congeneric species. *PLoS One*. 14(3):e0214162
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32(5):1792–1797.
- Jiang BS, Wen GX, Tang Y, Jiang QY. 2008. Effects of different treatments on cuttage cultivation and growth of *Keteleeria calcarea*. *Guihaia*. 28(4):549–552.
- Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW, Yi TS, Li DZ. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol*. 21(1):241
- Li H-T, Yi T-S, Gao L-M, Ma P-F, Zhang T, Yang J-B, Gitzendanner MA, Fritsch PW, Cai J, Luo Y, et al. 2019. Origin of angiosperms and the puzzle of the Jurassic gap. *Nat Plants*. 5(5):461–470.
- Nie L, Cui Y, Chen X, Xu Z, Sun W, Wang Y, Song J, Yao H. 2020. Complete chloroplast genome sequence of the medicinal plant *Arctiumlappa*. *Genome*. 63(1):53–60.
- Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, Yao H, Sun C, Li X, Li C, et al. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One*. 8(2):e57607
- Qu XJ, Moore MJ, Li DZ, Yi TS. 2019. PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods*. 15:50
- Sabater B. 2018. Evolution and function of the chloroplast. *Current investigations and perspectives*. *IJMS*. 19(10):3095.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30(9):1312–1313.
- 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.
- Wu CS, Lai YT, Lin CP, Wang YN, Chaw SM. 2009. Evolution of reduced and compact chloroplast genomes (cpDNAs) in gnetophytes: selection toward a lower-cost strategy. *Mol Phylogenet Evol*. 52(1):115–124.
- Xie WL, Chai SF, Jiang YS, Tang JM, Chen ZY, Zou R, Wei X. 2017. ISSR analysis on genetic diversity of *Keteleeria calcarea*. *Guihaia*. 37(01):36–41.
- Yi DK, Yang JC, So S, Joo M, Kim DK, Shin CH, Lee YM, Choi K. 2016. The complete plastid genome sequence of *Abies koreana* (Pinaceae: Abietoideae). *Mitochondrial DNA Part A*. 27(4):2351–2353.
- Zha X, Wang X, Li J, Gao F, Zhou Y. 2020. Complete chloroplast genome of *Sophora alopecuroides* (Papilionoideae): molecular structures, comparative genome analysis and phylogenetic analysis. *J Genet*. 99:13(2020).