

## The complete chloroplast genome sequence of *Thuja koraiensis* from Changbai Mountain in China

Cheng Jun Yang, Hang Zhang, Yi Liang Bai and Tian Tian Yang

Northeast Forestry University, Harbin, Heilongjiang, P. R. China

### ABSTRACT

In this study, the complete chloroplast genome of *Thuja koraiensis* was sequenced and analyzed. The complete chloroplast genome of *T. koraiensis* was 130,027 bp in length, encoding a total of 116 genes, including 80 protein-coding genes, 32 tRNAs, and 4 rRNAs, with a CG content of 34.24%. The phylogenetic analysis of *T. koraiensis* was carried out to determine the position of Thujoideae in the phylogenetic evolution.

### ARTICLE HISTORY

Received 23 December 2019  
Accepted 19 January 2020

### KEYWORDS

*Thuja koraiensis*; chloroplast genome; phylogenetic analysis; Thujoideae



*Thuja koraiensis* is an evergreen tree and belongs to the genus *Thuja*, and *T. koraiensis* is an endangered species. Its distribution area is limited, and its population is sparse. There are sporadic distributions only in mountains of 700–1400 m above sea level in Yanji, Changbai Mountain, Jilin, China, and northern Korea. *Thuja koraiensis* is included in the 'Red List of Species in China' as an endangered species of extremely small population, and is classified as a second-level key protected plant in the 'National List of Key Protected Wild Plants (First Batch)', a first-level key in Jilin Province Protecting plants is an urgently needed species in Changbai Mountains (Zhou 2006). It usually grows in moist air and humus-rich valleys and can grow on barren ridges and bare rocky seams. The wood of *T. koraiensis* is solid and durable, which can be used for construction, ships, furniture materials, and *T. koraiensis* is graceful, resistant to trimming, easy to shape, can be used as a landscaping tree species, and can also be used as a bonsai ornamental.

Chloroplasts are common organelles in higher plant cells, belonging to semi-autonomous organelles, with relatively independent genetic material, namely the chloroplast genome (Jing et al. 2018). Chloroplast genomes of most angiosperms are inherited from the maternal line, while only a few plants derive their chloroplast genomes from either parental or paternal lines (McCauley et al. 1996). Chloroplast genomes of gymnosperms are generally inherited from the father line (Soltis and Soltis 2000). Chloroplast genome sequence and structure are conservative, most of the plants of chloroplast genome in pairs link structure, contains four partitions, respectively, two reverse repeat area (IRA/IRB), a large single-copy area (LSC), and a small single-copy area (SSC), four parts and its gene length is relatively fixed, generally between 120 and 180 kbp (Liang et al. 2018). However, the reverse repeat

region of the chloroplast genome of gymnosperms is generally degraded seriously, with only a few dozen to a few hundred bp retained. Some species do not have reverse repeat regions, or there are more than one pair of reverse repeat regions (Luo et al. 2018). Such chloroplast genomes, containing the genetic information, are not as good as nuclear genomes, except for their stable structure, highly conserved copy number, easy to extract, and purification (Daniell et al. 2016). Therefore, it is of great significance in plant phylogeny, species identification, and genetic transformation. The complete chloroplast genome of *T. koraiensis* is sequenced and analyzed in this study in order to evaluate the phylogenetic position of *T. koraiensis* in the Thujoideae family. The complete chloroplast genome sequence of *T. koraiensis* has been deposited into the GenBank with the accession number MN820657.

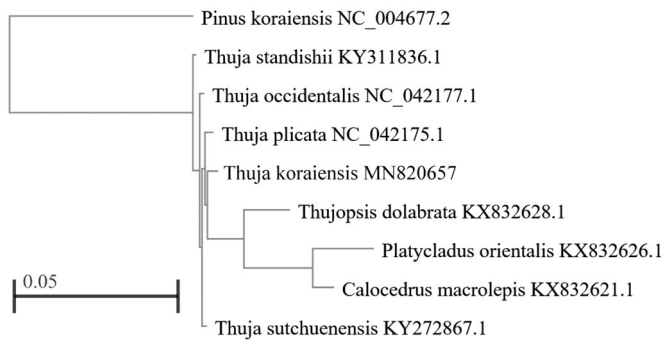
This experimental material was collected from the southern slope of Jilin Changbai Mountain National Nature Reserve (China, Jilin Province, 41°24'–41°47'N, 127°55'–128°11'E). The voucher specimen (NEFI20190701yang01) was deposited at Herbarium, Northeast Forestry University, Harbin, China.

After the fresh leaves are collected, they are immediately wrapped in tinfoil and placed in liquid nitrogen. The fresh leaves are then stored at  $-80^{\circ}\text{C}$ . This project uses the Whole Genome Shotgun (WGS) strategy to build a library of different inserts (Altschul et al. 1990) and uses second-generation sequencing technology (Next-Generation Sequencing, NGS), based on the HiSeq sequencing platform, paired-end (PE) testing of these libraries (Blake and Cohen 2001). The samples were subjected to DNA extraction, purification, database construction, and sequencing to obtain offline data. The offline data is saved in the Paired-end FASTQ format. The sequencing data contains some reads with adapters and low

**CONTACT** Tian Tian Yang  767031654@qq.com  Northeast Forestry University, No.26 Hexing Road, Harbin 150040, Heilongjiang, P. R. China

© 2020 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 inferred by maximum-likelihood (ML) method based on the complete chloroplast genome of 8 species of Thujoideae and taking *Pinus koraiensis* as an outgroup.

quality. These sequences will cause a great interference to subsequent information analysis. In order to ensure the quality of subsequent information analysis. The offline data is further filtered (Coil et al. 2015). Spades 3.12 was used to assemble the de-sequenced sequencing data from scratch to get the final spliced sequence (Schubert et al. 2016).

The chloroplast genome of *T. koraiensis* is 130,027 bp in length, and 116 genes are annotated, including 80 protein-coding genes, 32 tRNAs, 4 rRNAs, and CG% content of 34.24%. Among them *trnK-UUU*, *rpoC1*, *atpF*, *trnG-UCC*, *trnL-UAA*, *ndhA*, *trnA-UGC*, *trnI-GAU*, *ndhB*, *rpl2*, *rpl16*, *petD*, *petB*, and *trnV-UAC* contain a single intron. *Ycf3* and *rps12* contain two introns. The chloroplast genome structure of *T. koraiensis* is different from the four-segment structure of ordinary chloroplast plants. No inverted repeat region was found in the chloroplast genome of *T. koraiensis*.

Phylogenetic analysis was performed using the complete chloroplast genomes of *T. koraiensis* with 7 other species of Thujoideae reported in Genbank of NCBI and taking *Pinus koraiensis* as an outgroup database by the maximum-likelihood method in MEGA version 5.0 (Tamura et al. 2011). The results showed that among the subfamilies of Thujoideae, *T. koraiensis* has the closest kinship with *Calocedrus macrolepis*, *Platycladus orientalis*, and *Thujopsis dolabrata*, and it has the furthest relationship with *Thuja standishii* (Figure 1).

## Disclosure statement

The authors declare no conflicts of interest. The authors alone are responsible for the content and writing of this article.

## Funding

This work was supported by the Fundamental Research Funds for the Central Universities [2572019BA08] and Wildlife Protection and Management Project of the State Forestry Administration (Asexual breeding technology and wild tending of *T. koraiensis* in northeast China).

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.
- Blake JD, Cohen FE. 2001. Pairwise sequence alignment below the twilight zone. *J Mol Biol.* 307(2):721–735.
- Coil D, Jospin G, Darling AE. 2015. A5-MiSeq: an updated pipeline to assemble the naydin genomes from Illumina MiSeq data. *Bioinformatics.* 31(4):587–589.
- Daniell H, Lin CS, Ming Y, Chang W. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17(1):134–163.
- Jing M, Li XP, Li HT, Yang JB, Wang H, He J. 2018. Comparative analysis of the complete chloroplast genomes of four aconitum medicinal species. *Molecules.* 23(5):1015.
- Liang FP, Wen XG, Gao HY, Zhang Y. 2018. Analysis of chloroplast genomes features of Asteraceae species. *Genom Appl Biol.* 37(12): 5437–5447.
- Luo QF, Feng YH, Jia W, Chen H, Yang ZQ. 2018. Sequencing and characteristics analysis of chloroplast genome of *Pinus massoniana*. *Guangxi Forestry Sci.* 47(04):396–402.
- McCauley DE, Stevens JE, Peroni PA, Raveill J. 1996. The spatial distribution of chloroplast DNA and allozyme polymorphisms within a population of *Silene alba* (Caryophyllaceae). *Am J Bot.* 83(6):727–731.
- Schubert M, Lindgreen S, Orlando L. 2016. Adapter removal v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes.* 9(1):88.
- Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc Nat Acad Sci.* 97(13):7051–7057.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28(10):2731–2739.
- Zhou Q. 2006. Preliminary study of the evaluation system of precious and endangered wild officinal plants in Changbai Mountains. *Acta Botanica Boreali-Occidentalia Sinica.* 3:599–605.