MITOCHONDRIAL DNA PART B
2020, VOL. 5, NO. 1, 578–580
https://doi.org/10.1080/23802359.2019.1710587

Taylor & Francis
Taylor & Francis Group

RAPID COMMUNICATION

OPEN ACCESS Check for updates

High-throughput sequencing-based assembly of chloroplast genomes of five pine tree species

Qunfeng Luo, Zhangqi Yang, Yuanheng Feng, Jie Jia, Jianhui Tan and Peidong Yan

Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation, Guangxi Institute of Forestry Science, Nanning, China

ABSTRACT

Pinus plants are the largest existing group of gymnosperms and one of the most highly differentiated taxa. Due to its huge ecological, economic, and scientific value, the genetic diversity and the relationship between the intraspecific evolution of *Pinus* plants have gained wide attention. In this study, the chloroplast genomes of several common pine trees in southwest and south China, including *P. massoniana* (masson pine), *P. yunnanensis* (yunnan pine), *P. latteri* (south asia pine), *P. crassicorticea* (la ya pine), and *P. elliottii* (slash pine), and entire cpDNA sequences were obtained. Characteristics including the structure, repeated sequence, and codon bias of the cpDNA for these five pine tree species were analyzed.

ARTICLE HISTORY
Received 18 July 2019
Accepted 23 December 2019

KEYWORDS
Gymnosperm; chloroplast genome; codon bias; phylogenetic analysis

The *Pinus* is mainly distributed in the northern hemisphere and is divided into the subgenera *Pinus* and *Strobus*. Due to its huge ecological, economic, and scientific value, the genetic diversity and the relationship between the intraspecific evolution of *Pinus* plants have gained wide attention (Mardanov et al. 2008; Huang et al. 2014).

In this study, high-throughput sequencing technology was used to sequence the chloroplast genomes of several common pine trees, including *P. massoniana* (Gui GC833D), *P. yunnanensis* (Gui GC394E), *P. latteri* (Gui GC1382E), *P. crassicorticea* (Gui GC846E), and *P. elliottii* (Gui GC111E), with the GenBank accession codes: MH701846\MK135066\MK 000550\MK105898\MK105897. All samples were collected from the germplasm resource bank of the Nanning Forestry Research Institute (N23°10′, E108°00′) located in Wuming County, Guangxi, China. DNA library construction and sequencing were performed using the Illumina HiSeq PE150 strategy. Using the published chloroplast genome sequence of *P. taeda* (loblolly pine) as a reference, chloroplast reads for five samples were extracted and analyzed (Besemer et al. 2001; Lohse et al. 2007; Luo et al. 2012). Among the five chloroplast genomes, *P. elliottii* had the largest chloroplast genome, with a total length of 119,876 bp; *P. latteri* had the smallest chloroplast genome, with a total length of 119,715 bp. The difference in GC content was within 0.1% among the five pine tree species (Figure 1).

The coding gene prediction showed that the number of coding genes was 48 in *P. massoniana* cpDNA and in *P. crassicorticea* cpDNA, and 47 in *P. yunnanensis* cpDNA, *P. latteri* cpDNA, and *P. elliottii* cpDNA. *P. latteri* had the longest average gene length (954 bp), and *P. yunnanensis* had the shortest (826 bp). *P. massoniana*, and *P. crassicorticea* had the highest ratio of gene length to whole genome length (37.59%), and *P. yunnanensis* had the smallest ratio (32.39%). Forthese protein-coding chloroplast genes, base A had the highest frequency at the first position of corresponding codons, whereas base T had the highest frequency at the second and third positions. Thus, the codons of chloroplast proteins prefer to end with A–– and T––. Except for *P. yunnanensis*, more than 70% of the codons in the other four pine tree species ended in A–– or T––. Based on the RSCU (relative synonymous codon usage, Xu et al. 2010) values of the chloroplast protein-coding genes in the five pine tree species, the results showed that these species had an adequate preference for codon use.

The chloroplast genomes of 16 *Pinus* plant species were subjected to sequence alignment to construct phylogenetic trees for the study of genetic geography. MAFFT v7.394 software was first used for sequence alignment (Kazutaka and Standley 2013), and the ML (maximum-likelihood) and BI (Bayesian inference) methods were then used to construct phylogenetic trees. The results indicated that the phylogenetic trees constructed by the two methods were virtually consistent (Figure 2).

CONTACT Zhangqi Yang yangzhangqi@163.com Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation, Guangxi Institute of Forestry Science, Nanning 530002, China

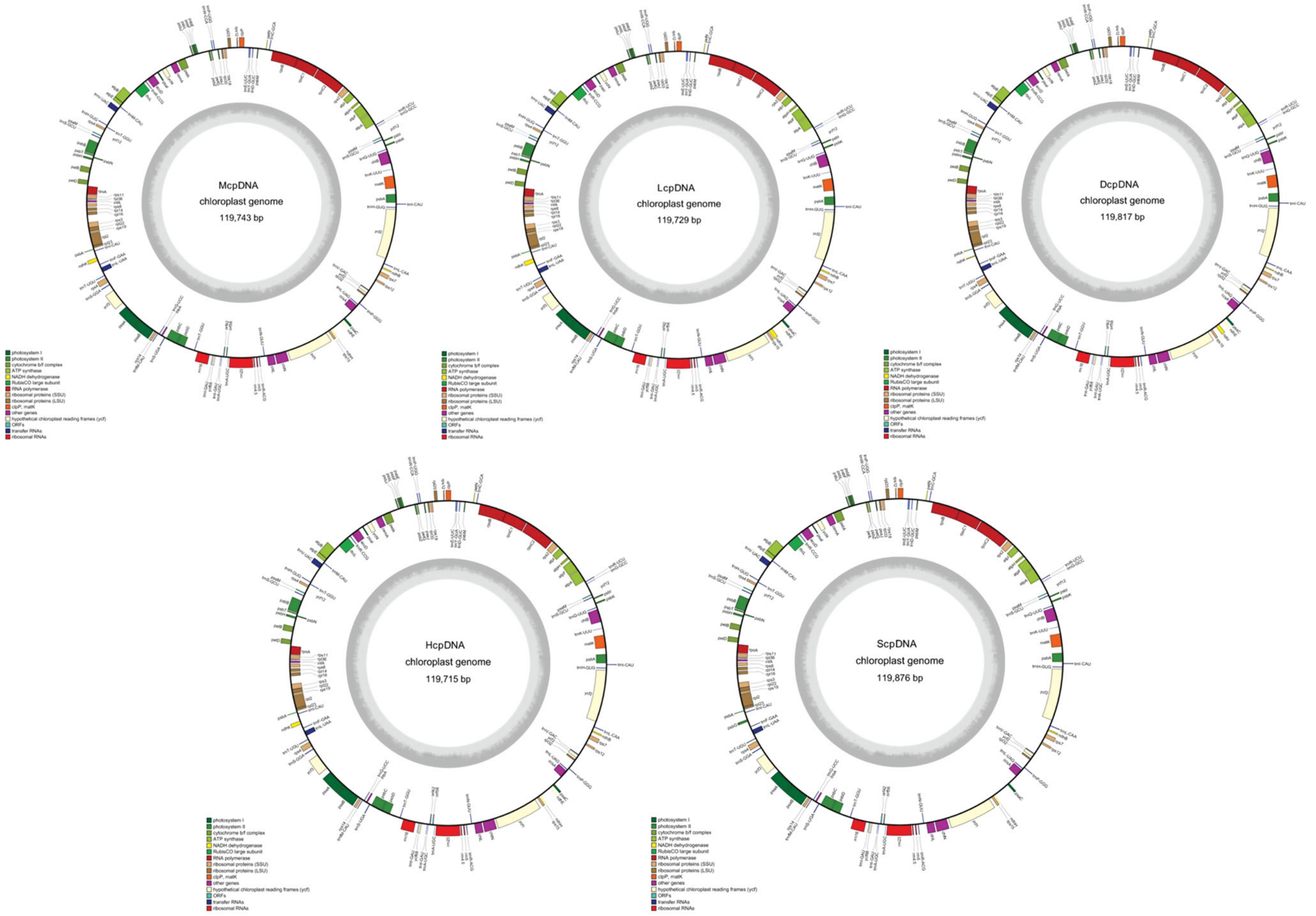

Figure 1. Gene maps of the chloroplast genomes.

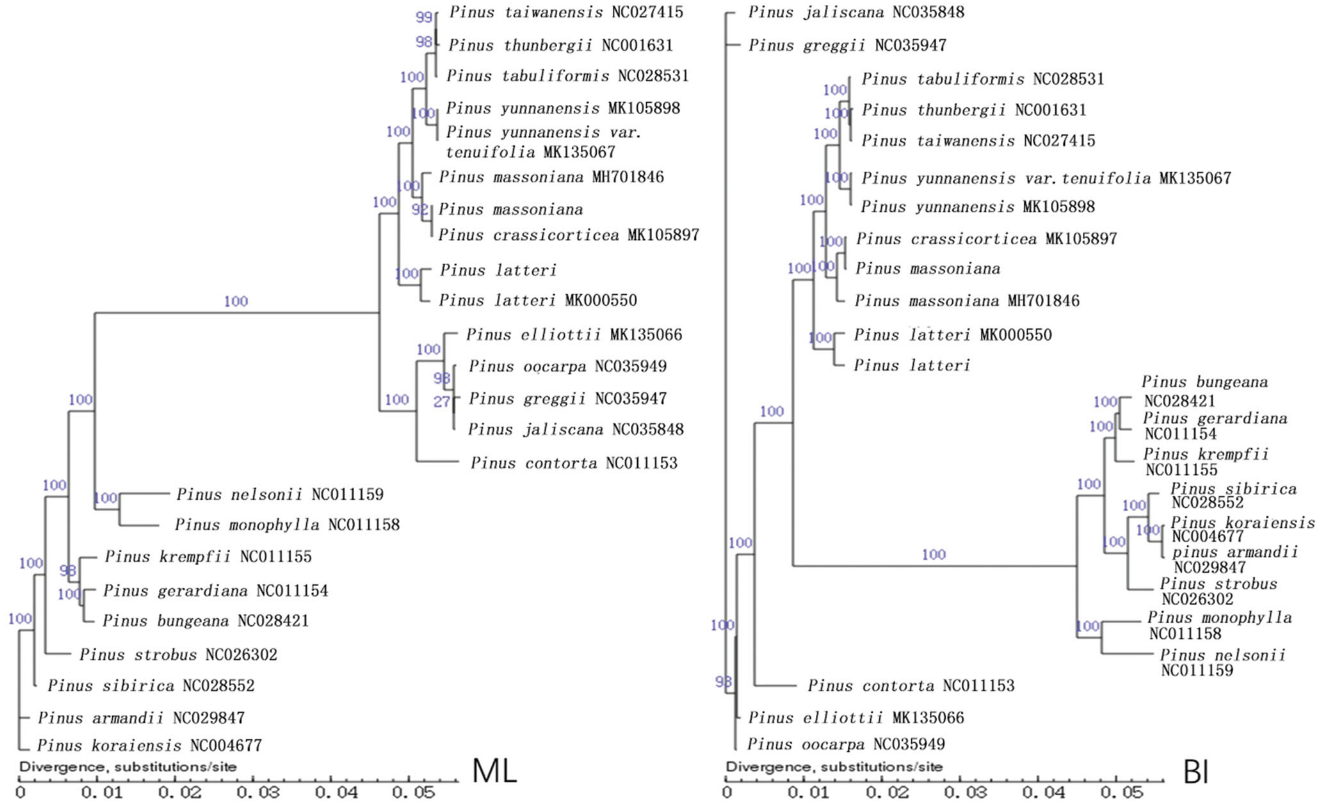
99 Pinus taiwanensis NC027415
98 Pinus thunbergii NC001631
100 Pinus tabuliformis NC028531
100 Pinus yunnanensis MK105898
100 Pinus yunnanensis var. tenuifolia MK135067
100 Pinus massoniana MH701846
100
100
92 Pinus massoniana
Pinus crassicorticea MK105897
100 Pinus latteri
Pinus latteri MK000550
100
100 Pinus elliottii MK135066
98 Pinus oocarpa NC035949
27 Pinus greggii NC035947
Pinus jaliscana NC035848
100
Pinus contorta NC011153
100
100
100 Pinus nelsonii NC011159
Pinus monophylla NC011158
100
98 Pinus krempfii NC011155
100 Pinus gerardiana NC011154
Pinus bungeana NC028421
100
Pinus strobus NC026302
Pinus sibirica NC028552
Pinus armandii NC029847
Pinus koraiensis NC004677
Divergence, substitutions/site
0 0.01 0.02 0.03 0.04 0.05
ML
Pinus jaliscana NC035848
Pinus greggii NC035947
100 Pinus tabuliformis NC028531
100 Pinus thunbergii NC001631
100 Pinus taiwanensis NC027415
100 Pinus yunnanensis var. tenuifolia MK135067
100 Pinus yunnanensis MK105898
100 Pinus crassicorticea MK105897
100 100 Pinus massoniana
Pinus massoniana MH701846
100 Pinus latteri MK000550
Pinus latteri
100 Pinus bungeana NC028421
100 Pinus gerardiana NC011154
Pinus krempfii NC011155
100
100 Pinus sibirica NC028552
100 Pinus koraiensis NC004677
100 pinus armandii NC029847
Pinus strobus NC026302
100 Pinus monophylla NC011158
Pinus nelsonii NC011159
100
100
100
98 Pinus contorta NC011153
Pinus elliottii MK135066
Pinus oocarpa NC035949
Divergence, substitutions/site
0 0.01 0.02 0.03 0.04 0.05
BI

Figure 2. Clustering of 24 *Pinus* species based on complete chloroplast genome sequences using the maximum-likelihood (ML) and Bayesian Inference likelihood (BI) methods.

Disclosure statement

The authors declare that they have no conflict of interest.

Funding

This study was supported by the National Natural Science Foundation of China [31560216], the National "Thirteenth Five-Year" Key R&D Project [2017YFD0600303], Guangxi Science and Technology Major Project [GuikeAA17204087-1], Guangxi Science and Technology Project [GuikeAD16380010] and Autonomous subjects of Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation [2019-A-03-03].

References

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29(12):2607–2618.

Huang H, Shi C, Liu Y, Mao SY, Gao LZ. 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationshi*ps*. BMC Evol Biol. 14(1):151.

Kazutaka K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Lohse M, Drechsel O, Bock R. 2007. Organellar Genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 52(5–6):267–274.

Luo R, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GZ, Chen YX, Pan Q, Liu YJ, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaSci. 1(1):6.

Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kolganova TV, Skyabin KG. 2008. Complete sequence of the Duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. J Mol Evol. 66(6):555–564.

Xu C, Ben AL, Cai XN. 2010. Analysis of synonymous codon usage in chloroplast genome of *Phapaenopsis aphrodite* subsp. Formosana. Mol Plant Breed. 8:945–950.