


The complete chloroplast genome sequence of *Santalum album*

Dejun Yang^{a,b}, Qiong Qiu^a, Linhong Xu^a, Yumei Xu^a and Yi Wang^{a,b} 

^aInstitute of Tropical Forestry, Yunnan Academy of Forestry, Puwen, People's Republic of China; ^bLaboratory of Forest Plant Cultivation and Utilization, Yunnan Academy of Forestry, Kunming, People's Republic of China

ABSTRACT

The first complete chloroplast genome (cpDNA) sequence of *Santalum album* was determined from Illumina HiSeq pair-end sequencing data in this study. The cpDNA is 144,101 bp in length, contains a large single copy region (LSC) of 83,796 bp and a small single copy region (SSC) of 11,277 bp, which were separated by a pair of inverted repeats (IR) regions of 24,514 bp. The genome contains 123 genes, including 80 protein-coding genes, 8 ribosomal RNA genes, and 35 transfer RNA genes. The overall GC content of the whole genome is 38.0%, and the corresponding values of the LSC, SSC, and IR regions are 35.9%, 31.4%, and 43.1%, respectively. Further phylogenomic analysis showed that *S. album* and *Osyris alba* clustered in a clade in Santalales order.

ARTICLE HISTORY

Received 29 November 2019
Accepted 7 December 2019

KEYWORDS

Santalum album;
chloroplast; Illumina
sequencing;
phylogenetic analysis

Santalum album is the species within the family Santalaceae. It is widely distributed in India, Malaysia, and Australia, and commonly known as sandalwood (Kim et al. 2005). The essential oil of sandalwood widely used in perfumes, cosmetics, and sacred unguents (Jones et al. 2006). Sandalwood oil has various biological activities, such as antiviral, antibacterial

(Benencia and Courreges 1999) and antitumor activities (Kim et al. 2006). Sandalwood is also used in treatment of ailments like vomiting, poisoning, eye infections and other diseases (Sindhu et al. 2010). Therefore, *S. album* has huge medicinal value. However, there have been no chloroplast genomic studies on *S. album*.

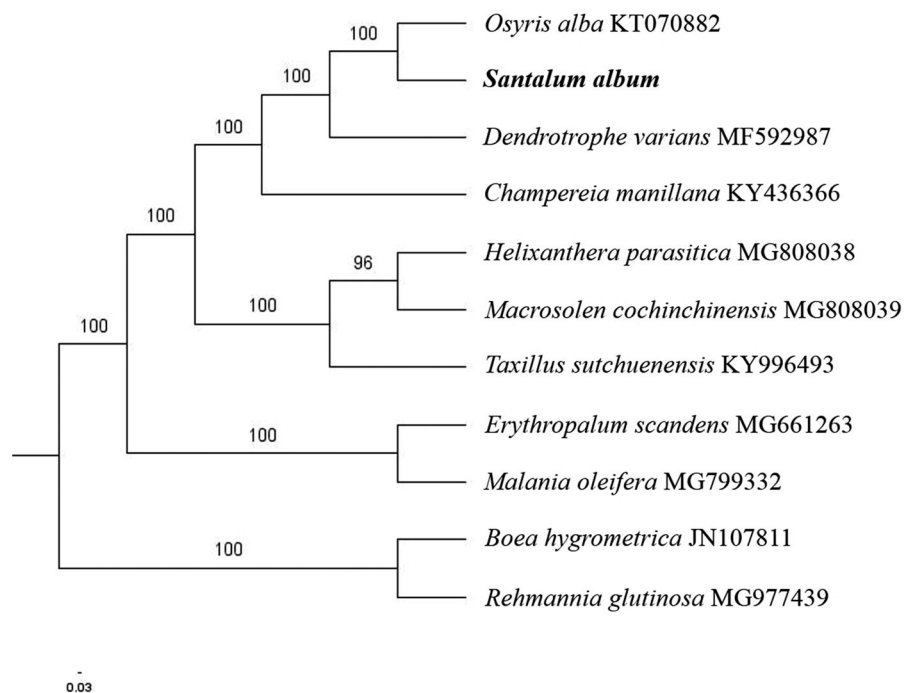


Figure 1. The maximum-likelihood tree based on the nine chloroplast genomes of order Santalales. The bootstrap value based on 1000 replicates is shown on each node.

Herein, we reported and characterized the complete *S. album* plastid genome (MN106256). One *S. album* individual (specimen number: 201807061) was collected from Puwen, Yunnan Province of China (22°25'39"N, 101°6'46"E). The specimen is stored at Yunnan Academy of Forestry Herbarium, Kunming, China and the accession number is YAFH0012863. DNA was extracted from its fresh leaves using DNA Plantzol Reagent (Invitrogen, Carlsbad, CA).

Paired-end reads were sequenced by using Illumina HiSeq system (Illumina, San Diego, CA). In total, about 31.1 million high-quality clean reads were generated with adaptors trimmed. Aligning, assembly, and annotation were conducted by CLC de novo assembler (CLC Bio, Aarhus, Denmark), BLAST, GeSeq (Tillich et al. 2017), and GENEIOUS v 11.0.5 (Biomatters Ltd, Auckland, New Zealand). To confirm the phylogenetic position of *S. album*, other eight species of order *Santalales* from NCBI were aligned using MAFFT v.7 (Kato and Standley 2013). The Auto algorithm in the MAFFT alignment software was used to align the eleven complete genome sequences and the G-INS-i algorithm was used to align the partial complex sequences. The maximum-likelihood (ML) bootstrap analysis was conducted using RAxML (Stamatakis 2006); bootstrap probability values were calculated from 1000 replicates. *Boea hygrometrica* (JN107811) and *Rehmannia glutinosa* (MG977439) were served as the out-group.

The complete *S. album* plastid genome is a circular DNA molecule with the length of 144,101 bp, contains a large single copy region (LSC) of 83,796 bp and a small single copy region (SSC) of 11,277 bp, which were separated by a pair of inverted repeats (IR) regions of 24,514 bp. The overall GC content of the whole genome is 38.0%, and the corresponding values of the LSC, SSC, and IR regions are 35.9%, 31.4%, and 43.1%, respectively. The plastid genome contained 123 genes, including 80 protein-coding genes, 8 ribosomal RNA genes, and 35 transfer RNA genes. Phylogenetic analysis showed that *S. album* and *Osyris alba* clustered in a unique clade in *Santalales* order (Figure 1). The determination of the complete plastid genome sequences provided new molecular data to illuminate the order *Santalales* evolution.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Yunnan Key Research and Development Project in Forestry [2018BB007] and Construction Project of Xishuangbanna High-Efficiency Cultivation Test and Demonstration Base for Valuable Timber Tree Plantation.

ORCID

Yi Wang  <http://orcid.org/0000-0003-3089-8184>

References

- Benencia F, Courreges MC. 1999. Antiviral activity of sandalwood oil against herpes simplex viruses-1 and-2. *Phytomedicine*. 6(2): 119–123.
- Jones CG, Ghisalberti EL, Plummer JA, Barbour EL. 2006. Quantitative co-occurrence of sesquiterpenes; a tool for elucidating their biosynthesis in Indian sandalwood, *Santalum album*. *Phytochemistry*. 67(22): 2463–2468.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Kim TH, Ito H, Hatano T, Takayasu J, Tokuda H, Nishino H, Machiguchi T, Yoshida T. 2006. New antitumor sesquiterpenoids from *Santalum album* of Indian origin. *Tetrahedron*. 62(29):6981–6989.
- Kim TH, Ito H, Hayashi K, Hasegawa T, Machiguchi T, Yoshida T. 2005. Aromatic constituents from the heartwood of *Santalum album* L. *Chem Pharm Bull*. 53(6):641–644.
- Sindhu RK, Upma KA, Arora S. 2010. *Santalum album* linn: a review on morphology, phytochemistry and pharmacological aspects. *Int J PharmTech Res*. 2(1):914–919.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 22(21):2688–2690.
- 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.