

The complete chloroplast genome sequence of *Nicotiana debneyi* (Solanaceae)

Jianmin Zeng^a , Changjun Huang^a, Yong Liu^a, Cheng Yuan^a, Haiqin Yu^a and Biqing Song^b

^aYunnan Academy of Tobacco Agricultural Sciences, Kunming, P.R. China; ^bYuxi Zhongyan Tobacco Seed Co., Ltd, Yuxi, P.R. China

ABSTRACT

In present study, we report the complete chloroplast genome of *Nicotiana debneyi*, a species endemic to eastern coast of Australia. The total genome size of *N. debneyi* is 156,073 bp in length, containing a large single-copy region of 86,672 bp, a small single-copy region of 18,581 bp, and two inverted repeat regions of 25,410 bp. The all GC content of *N. debneyi* chloroplast genome is 38.4%. It encodes a total of 129 unique genes, including 84 protein-coding genes, 37 tRNA genes, and eight rRNA genes, of which seven tRNA, four rRNA and seven protein-coding genes are duplicated in the IR. Sixteen genes contain a single intron, and only two genes have two introns. Phylogenetic analysis results strongly supported that *N. debneyi* was closely related to *Nicotiana sylvestris* and *Nicotiana tabacum*.

ARTICLE HISTORY

Received 10 November 2020
Accepted 24 February 2021

KEYWORDS

Chloroplast genome;
Nicotiana debneyi;
phylogenetic analysis

Nicotiana has been used as an important model plant in life science research, from around evolution, cytogenetics, interspecific hybridization and medical drug development (Winter 2000). It contains more than eighty species around the world, mainly distributed in the Americas, Australia and some species in Africa. *Nicotiana debneyi* is a tetraploid ($2n = 4x = 48$) wild tobacco native to the eastern coast of Australia classified in the section *Suaveolentes*. Resistances to black root rot, and blue mold and powdery mildew have been introduced from this species into *Nicotiana tabacum*, and the cross-compatible with *N. tabacum* is cytoplasmic male-sterile (Japan Tobacco Inc 1994).



The chloroplast genome sequence has unique characteristics, such as uniparental inheritance, conserved sequence composition in a typical quadripartite genome structure, numerous variable sites, and coding regions (Bi et al. 2018). Chloroplast DNA sequence data are a versatile tool for plant identification or barcoding and to enhance our understanding about phylogeny of plant species. Studies on the chloroplast genome could also contribute useful information for our understanding of the molecular biology, physiology, and evolutionary of chloroplasts. However, the plastid genome of *N. debneyi* has not been reported. Herein, we report the complete chloroplast sequence of *N. debneyi* based on the next-generation sequencing and the annotated genomic sequence was deposited in the GenBank with the accession number MT985319, in order to provide resources for further genomic studies and chloroplast engineering.

Extraction of high-quality DNA was performed from 10g of fresh leaves collected in the tobacco research station of Yunnan Academy of Tobacco Agricultural Sciences, Yuxi City, Yunnan province, China (N24°21'25.75"; E102°32'37.39", voucher TW37), according to CTAB method (Doyle and Doyle

1987). Illumina 2×150 bp paired-end libraries were constructed with the NexteraXT DNA Library Preparation Kit (Illumina Inc., San Diego, CA) followed by sequencing on a Illumina Novaseq 6000 platform (Total Genomics Solution Limited, SZHT). Raw reads were filtered with Trimmomatic (Bolger et al. 2014), and de novo assembly of the plastid genome was carried out with the SPAdes-3.11.0 (<http://cab.spbu.ru/software/spades/>), using a k-mer set of 93,95,97,103,105,107,115 (Bankevich et al. 2012). The assembled chloroplast genome was annotated by Geneious v 9.0.2 (Kearse et al. 2012), followed by manual adjustments with the reference genome of *Nicotiana sylvestris* (GenBank: BK010741).

The complete chloroplast genome of *N. debneyi* (GenBank:MT985319) has a total length of 156,073 bp, which contains a pair of inverted repeats separated by small and large single copies. The genome consists of 129 genes and has 38.4% overall GC content. The genes can be classified into 37 tRNA, eight rRNA and 84 protein-coding genes, of which seven tRNA, four rRNA and seven protein-coding genes are duplicated in the IR. The four junctions between the single-copy segments and the inverted repeats were confirmed using PCR-based product sequencing. The results showed that the preliminary chloroplast genome was correctly assembled.

Twenty-five plastomes were aligned by MAFFT v.7.471 (Katoh et al. 2019) with default parameters. The best scoring Maximum Likelihood phylogenetic tree was calculated under GTR-GAMMA with 1000 bootstrap replicates using the program RAxMLv8.0 (Stamatakis 2014), including 20 representative species of Solanaceae and five outgroup species (Figure 1). The phylogenetic tree strongly indicated that *N. debneyi* formed a monophyletic clade and was closely related to two species (*Nicotiana sylvestris* and *Nicotiana tabacum*) by a high bootstrap value. The complete plastome

CONTACT Biqing Song  songbq@126.com  Yuxi Zhongyan Tobacco Seed Co., Ltd, Yuxi, Yunnan province, P.R. 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.

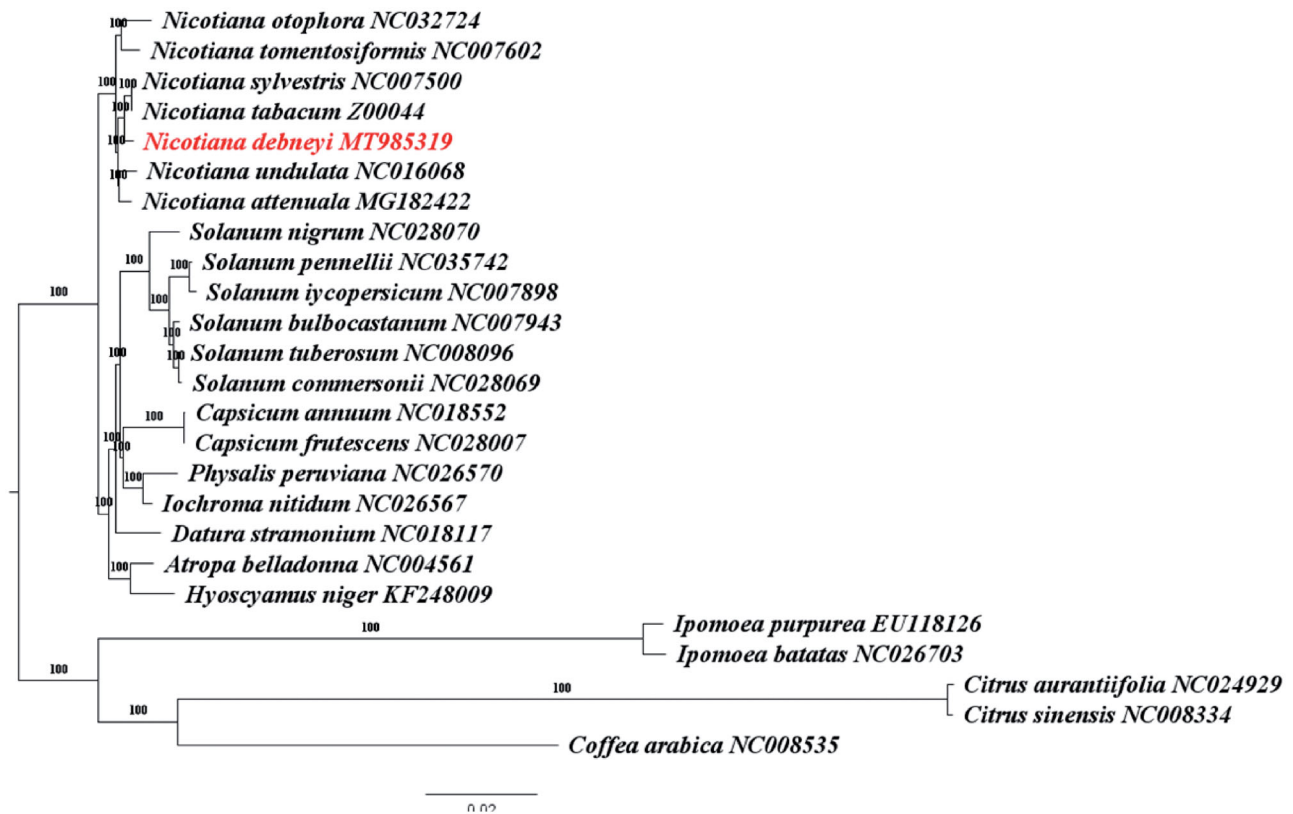


Figure 1. Phylogenetic relationships among 25 complete chloroplast genomes including 20 representative species of *Solanaceae* and five outgroup species. Bootstrap support values are given at the nodes.

can be subsequently utilized for genetic diversity and taxonomic studies of the genus *Nicotiana*.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the Grants from YNTC [2017YN04, 2019530000241004].

ORCID

Jianmin Zeng  <http://orcid.org/0000-0002-0989-5627>

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. MT985319. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA690194, SUB8850676, and SAMN17245913 respectively.

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Bi Y, Zhang MF, Xue J, Dong R, Du YP, Zhang XH. 2018. Chloroplast genomic resources for phylogeny and DNA barcoding: a case study on *Fritillaria*. *Sci Rep.* 8(1):1184
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120.
- Doyle JJ, Doyle JL. 1987. A Rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Japan Tobacco Inc. 1994. *The genus nicotiana illustrated*. Tokyo: Japan Tobacco, Inc.
- Katoh K, Rozewicki J, Yamada KD. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 20(4):1160–1166.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 28(12): 1647–1649.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9): 1312–1313.
- Winter JC. 2000. *Tobacco use by native North Americans: sacred smoke and silent killer*. Norman: University of Oklahoma Press.