

MITOGENOME ANNOUNCEMENT



Assembly and phylogenetic analysis of complete chloroplast genome of a wild tobacco *Nicotiana Attenuata*

Haifeng Lin^a and Di Bai^b

^aCollege of Information Science and Technology, Nanjing Forestry University, Nanjing, Jiangsu, China; ^bCollege of Engineering, Nanjing Agricultural University, Nanjing, Jiangsu, China

ABSTRACT

Nicotiana attenuata (*N. attenuata*) is an annual herb that induces nicotine, volatiles and phenolics in response to damage. In this study, we assembled the complete chloroplast (cp) genome of *N. attenuata* with the length of 155,886 bp, which was composed of a large single copy region (LSC) of 86,602 bp, a small copy (SSC) region of 18,518 bp and a pair of inverted repeats (IR) of 25,383 bp. This typical cp genome contains 112 unique genes, including 78 protein-coding genes, 30 tRNA genes and four rRNA genes. Phylogenetic analysis of 32 plant cp genomes based on 76 conserved protein-coding genes shows that *N. attenuata* is evolutionarily close to *N. tabacum*, then *Capsicum annuum*. The complete cp genome sequence of *N. attenuata* may provide a crucial foundation for the study of evolution, molecular biology and genetic engineering in *Nicotiana* species.

ARTICLE HISTORY

Received 21 September 2017
Accepted 26 October 2017

KEYWORDS

Nicotiana attenuata;
chloroplast genome; wild
tobacco; phylogeny

Nicotiana attenuata is a species of wild tobacco known locally by coyote tobacco, which is native to western North America and grows in many types of habitat. It is a glandular and sparsely hairy annual herb that can grow tall and compact or shrub-like to several feet tall and wide. *Nicotiana attenuata* can produce nicotine, a potent alkaloid that affects muscles, to master itself. In the case of *N. attenuata*, the concentration of nicotine can be up to three times the value found in common tobacco, *N. tabacum*, and it is so efficient in paralyzing or killing insects that it was widely used as an insecticide in the past (Baldwin et al. 1994).

Chloroplast is the green plastid observed in land plants, algae and some protists. The main role of chloroplasts is to conduct photosynthesis, while they also carry out lots of other functions, including fatty acid synthesis, much amino acid synthesis, and the immune response in plants. Generally, plant cp genomes are found to be highly conserved in gene order, content and organization (Henry and Henry 2005).

The leaf material from one single rosette-stage *N. attenuata* was harvested to DNA extraction in Washington County, UT (geographic coordinate: 37°16'48"N, 113°31'12"W) (Xu et al. 2017). The genomic DNA sequenced by Roche 454 and Illumina HiSeq2000 technologies was isolated with the CTAB-method (Bubner et al. 2004), and then deposited in Max Planck Institute for Chemical Ecology, 07745 Jena, Germany. The complete cp genome of *N. attenuata* was assembled using Newbler 3.0 (Wu et al. 2017), and corrected using BWA and SAMtools (Bi et al. 2016). The total length of the cp

genome was 155,886 bp and then submitted to GenBank with accession number MF577082.

The complete cp genome of *N. attenuata* exhibits a typical quadripartite structure of a LSC region of 86,602 bp, a SSC region of 18,518 bp and a pair of IR region of 25,383 bp. The overall GC content of this cp genome was 37.87%, the GC content of the LSC regions was 35.97%, the IR region was 43.20%, and the SSC region was 32.06%. Using the online annotation program DOGMA (Wyman et al. 2004), a total of 129 functional genes were predicted in the *N. attenuata* cp genome, including 112 unique genes, and 17 genes were found to have two copies. These unique genes were composed of 78 protein-coding genes, 30 tRNA genes and four rRNA genes. The multi-copied genes included six protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7*, and *ycf2*), seven tRNA genes (*trnA*-UGC, *trnI*-CAU, *trnI*-GAU, *trnL*-CAA, *trnN*-GUU, *trnR*-ACG, and *trnV*-GAC), and four rRNA genes (*rrn4.5*, *rrn5*, *rrn16*, and *rrn23*). The majority of the 129 genes in the *N. attenuata* cp genome did not contain introns, whereas 18 genes contained one intron (11 protein-coding and seven tRNA genes) and two (*ycf3*, *clpP*) contained two introns. Additionally, the *rps12* gene was trans-spliced, whose exon at the 5' end was located in the LSC region and another multi-copied exon at the 3' end was in the IR region. In order to confirm the phylogenetic position of *N. attenuata* in plants, a total of 76 protein-coding genes from 32 plant cp genomes were selected to reconstruct the neighbour-joining phylogenetic tree using MEGA6 (Tamura et al. 2013). As illustrated in Figure 1, the phylogenetic tree strongly support that the cp

CONTACT Di Bai  baidi000@njau.edu.cn  College of Engineering, Nanjing Agricultural University, Nanjing, Jiangsu, China

© 2017 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.

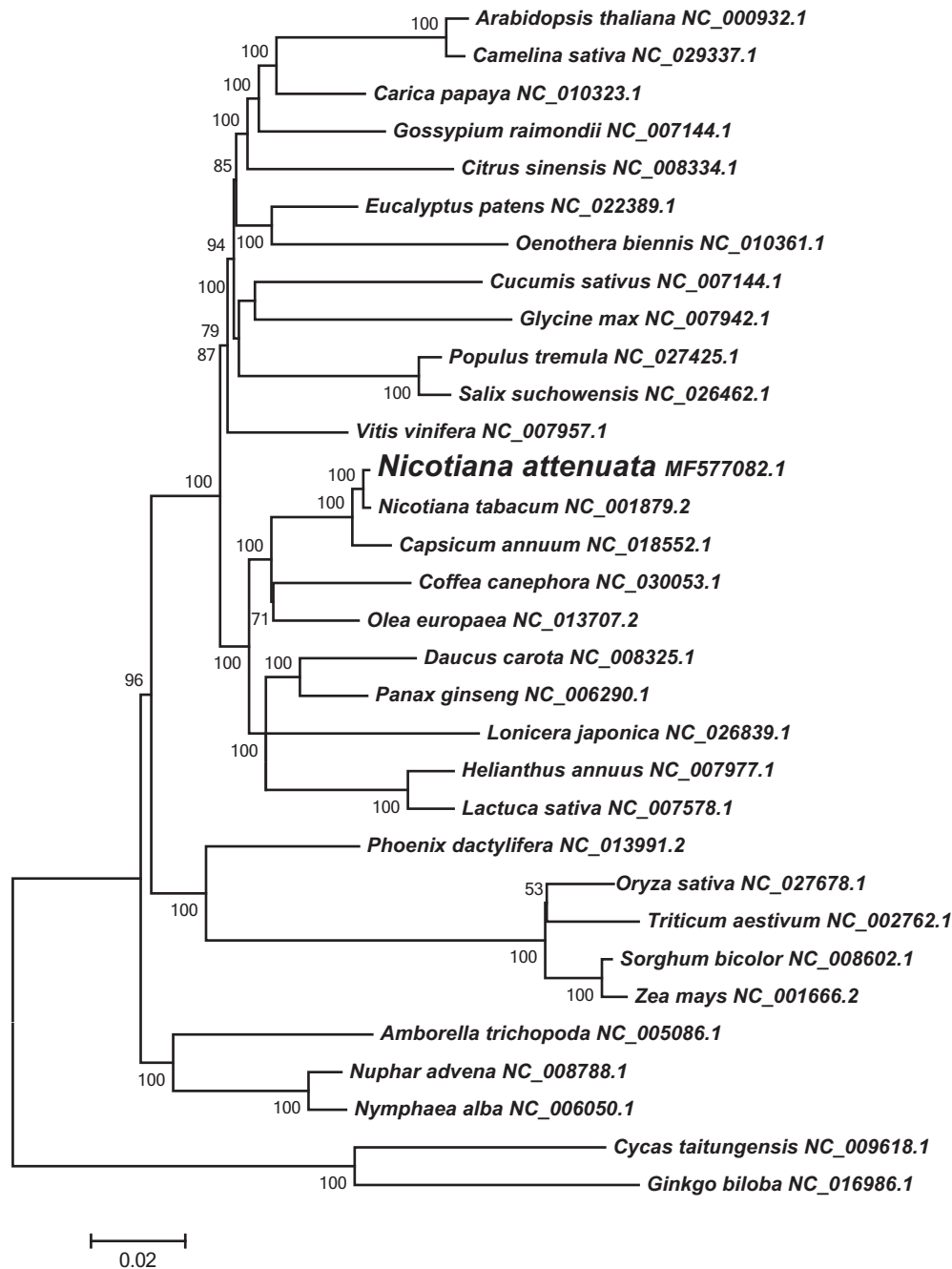


Figure 1. The neighbour-joining tree was constructed based on 76 protein-coding genes from 32 plant species. These protein-coding genes were aligned with MUSCLE and the phylogenetic tree was constructed using the neighbour-joining method in MEGA6. Numbers at the nodes are bootstrap support values.

genome of *N. attenuata* is evolutionarily closely related to another tobacco species *Nicotiana tabacum*, then *Capsicum annuum*.

Acknowledgements

The authors would like to acknowledge the Max Planck Institute for Chemical Ecology for their assistance in sampling the material of *Nicotiana attenuata*.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Funding

This work was supported by the Natural Science Foundation of Jiangsu Province under Grant no. BK20141474, Nanjing Forestry University Science and Technology Innovation Fund (Grant no. CX2016024), and The Jiangsu Overseas Research & Training Program for University Prominent Young & Middle-aged Teachers and Presidents.

References

- Baldwin IT, Staszak-Kozinski L, Davidson R. 1994. Up in smoke: I. Smoke-derived germination cues for postfire annual, *Nicotiana attenuata* torr. Ex. Watson. J Chem Ecol. 20:2345.
- Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N. 2016. Analysis of the complete mitochondrial genome sequence of the

- diploid cotton *Gossypium raimondii* by comparative genomics approaches. *BioMed Res Int*. 2016:5040598.
- Bubner B, Gase K, Baldwin IT. 2004. Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. *Bmc Biotechnol*. 4:14.
- Henry RJ, Henry RJ. 2005. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. *Biologist*. 41:123–124..
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.
- Wu D, Bi C, Wang X, Xu Y, Ye Q, Ye N. 2017. The complete chloroplast genome sequence of an economic plant *Coffea canephora*. *Mitochondrial DNA Part B*. 2:483–485.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*. 20:3252–3255.
- Xu S, Brockmüller T, Navarro-Quezada A, Kuhl H, Gase K, Ling Z, Zhou W, Kreitzer C, Stanke M, Tang H, et al. 2017. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci USA*. 114:6133–6138.