Complete chloroplast genome sequence of Coyote tobacco (Nicotiana attenuata, Solanaceae)

Péter Poczai, Ali Amiryousefi & Jaakko Hyvönen

Published online: 06 Nov 2017.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

MITOGENOME ANNOUNCEMENT

🔓 OPEN ACCESS Check for updates

Complete chloroplast genome sequence of Coyote tobacco (*Nicotiana attenuata*, Solanaceae)

Péter Poczai[a] 🔟 , Ali Amiryousefi[a,b] and Jaakko Hyvönen[b] 🔟

[a]Finnish Museum of Natural History (Botany), University of Helsinki, Helsinki, Finland; [b]Department Bioscience (Plant Biology), Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland

ABSTRACT

In this study, we announce the complete chloroplast genome sequence of *Nicotiana attenuata*. The genome sequence of 155,941 bp consists of two inverted repeat (IRa and IRb) regions of 25,438 bp each, a large single-copy (LSC) region of 86,513 bp and a small single-copy (SSC) region of 18,524 bp. The overall GC content is 37.9% and the GC contents of LSC, IRs, and SSC are 36%, 43.2%, and 32.1%, respectively. The plastome with 129 annotated unique genes includes 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Using the whole chloroplast genome sequences alignment of 16 Solanaceae species a phylogenetic hypothesis is presented validating the position of *N. attenuata* within Nicotianeae.
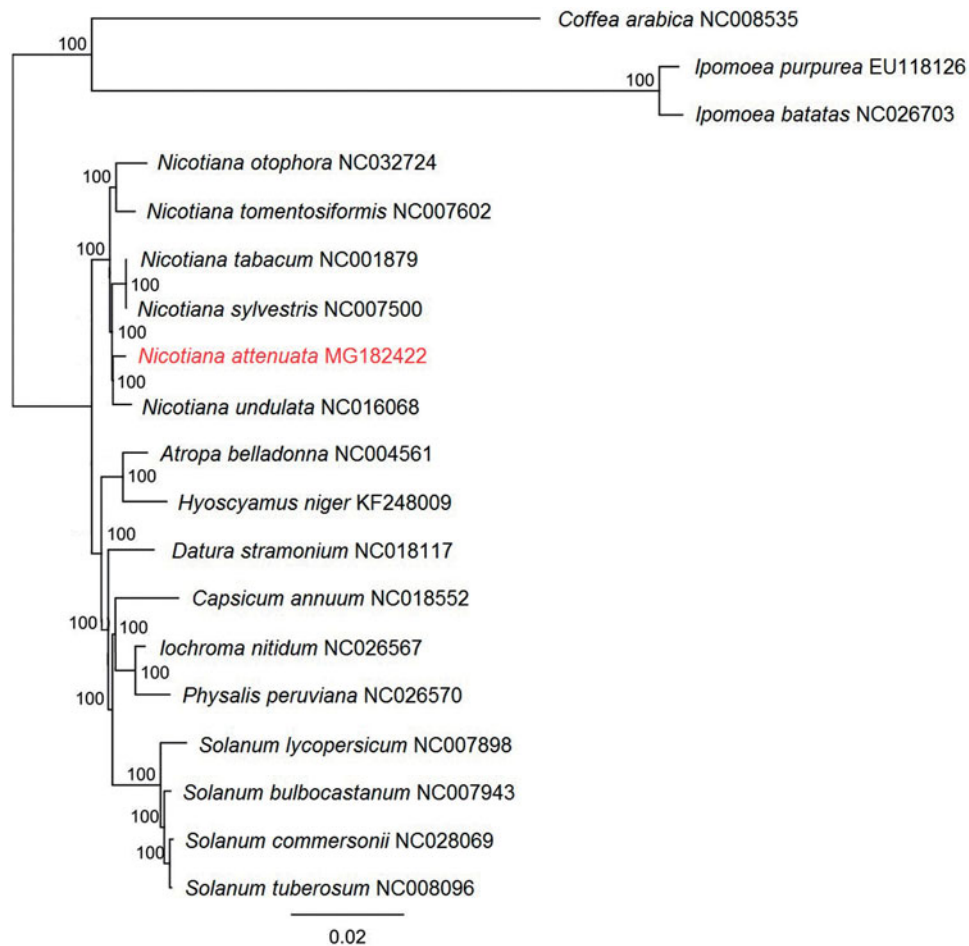
Nicotiana attenuata Torr. ex S.Watson is a diploid $(2n = 2x = 24)$ wild tobacco native to the Great Basin region of the United States. This plant has adapted to an ecological niche defined by the post-fire environment, where soils tend to be nitrogen-rich and biotic stresses are highly dynamic (Nelson et al. 2012). During the last decade, large amounts of genomic, transcriptomic and metabolomic data have been generated for this plant which has provided new insights into how native plants interact with herbivores, pollinators and microbes (Brockmöller et al. 2017). Despite the availability of such large amounts of omics data the plastid genome of coyote tobacco has not been assembled. However, studies on the chloroplast genome could contribute greatly to our understanding of the molecular biology, physiology, and biochemistry of chloroplasts. Chloroplast DNA sequence data are a versatile tool for plant identification or barcoding and to enhance our understanding about phylogeny of plant species. Herein, we report the complete chloroplast sequence of *N. attenuata* to provide resources for further (phylo)genomic studies and chloroplast engineering.

Extraction of DNA was performed from 20 g of fresh leaves according to Shi et al. (2012) collected in Kaisaniemi Botanical Garden, University of Helsinki, Finland (60.1753; 24.9460; voucher P.Poczai 0012887). Illumina 2×150 bp paired-end libraries were prepared with the TruSeq DNA Sample kit followed by sequencing on a MiSeq platform

(Illumina Inc., San Diego, CA). Raw reads were filtered with Trimmomatic (Bolger et al. 2014), and *de novo* assembly of the plastid genome was carried out with the Geneious R9.1.7 assembler platform (Kearse et al. 2012). We annotated the genome using Geneious and in-house scripts

The complete chloroplast genome of *N. attenuata* (GenBank accession no. MG182422) has a total length of 155,914 bp which is divided by two IR regions of 25,438 bp. The genome consists of 129 genes and has 37.9% overall GC content. The genes can be classified into 37 tRNA, eight rRNA and 84 protein-coding genes, from which seven tRNA, four rRNA and six protein-coding genes are duplicated in the IR.

Using the RAxMLv8.0 (Stamatakis 2014), we calculated the best scoring ML tree with 10,000 bootstrap replicates under GTR-GAMMA after running jModelTest2 (Darriba et al. 2012) including 15 representative species of Solanaceae and three outgroup species of Gentianales (Figure 1). The same matrix was analyzed also with parsimony as an optimality criterion using WinClada (Nixon 2002) and TNT (Goloboff et al. 2008). Prior to the analysis we used the WinClada command 'Mop uninformative characters' to exclude parsimony uninformative characters. This resulted in a matrix with 17,620 characters and due to its small size we were able to perform analyses using implicit enumeration of the TNT that ensures finding optimal tree(s).

CONTACT Péter Poczai ✉ peter.poczai@gmail.com 🏛 Finnish Museum of Natural History (Botany), University of Helsinki, PO Box 7, Helsinki FI-00014, Finland

Figure 1. The ML tree of 18 selected chloroplast genome sequences and *Nicotiana attenuata*. The values on the node show the bootstraps of 10,000 replicates and scale is substitution per site.

We expect this sequence to help further taxonomic studies of the genus *Nicotiana* and provide additional genomic resources for chloroplast engineering.

Acknowledgements

Disclosure statement

The authors claim no conflict of interest.

ORCID

Péter Poczai http://orcid.org/0000-0002-0107-1068
Jaakko Hyvönen http://orcid.org/0000-0001-7559-8295

References

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Brockmöller T, Ling Z, Li D, Gaquerel E, Baldwin IT, Xu S. 2017. *Nicotiana attenuata* Data Hub (NaDH): an integrative platform for exploring genomic, transcriptomic and metabolomic data in wild tobacco. BMC Genomics. 18:79.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and high-performance computing. Nature Math. 9:772.

Goloboff PA, Farris JS, Nixon KC. 2008. TNT, a free program for phylogenetic analysis. Cladistics. 24:774–786.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinfo. 28:1647–1649.

Nelson DC, Flematti GR, Ghisalberti EL, Dixon KW, Smith SM. 2012. Regulation of seed germination and seedling growth by chemical signals from burning vegetation. Annu Rev Plant Biol. 63: 107–130.

Nixon KC. 2002. WinClada 1.00.08. Available from: www.cladistics.com.

Shi C, Hu N, Huang H, Gao J, Zhao Y-J, Gao L-Z. 2012. An improved chloroplast DNA extraction procedure for whole plastid genome sequencing. PLoS One. 7:e31468.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.