

## The complete chloroplast genome of *Ocimum gratissimum* from India – a medicinal plant in the Lamiaceae

Raju Balaji, Kumar Ravichandiran, Tanuja and Madasamy Parani

Center for DNA Barcoding, Department of Genetic Engineering, SRM Institute of Science and Technology, Kattankulathur, India

### ABSTRACT

*Ocimum gratissimum* L. is an important medicinal species with several therapeutic applications. It is used in traditional medicine as a single drug and in formulations. We generated the complete chloroplast genome sequence of *O. gratissimum* by using Illumina paired-end sequencing data. The *O. gratissimum* chloroplast genome is 152,469 bp in length, containing a large single copy (LSC) region of 83,614 bp and a small single copy region (SSC) of 17,607 bp, separated by a pair of inverted repeats (IRs) of 25,624 bp. The genome contains 138 unique genes, including 85 protein-coding, 45 tRNA, and eight rRNA genes. Among them, six genes have one intron each, and two genes contain two introns. The overall GC content of the chloroplast genome is 37.8%, while the corresponding values of LSC, SSC, and IR regions are 35.6%, 31.7%, and 43.2%, respectively. Phylogenetic analysis with the complete chloroplast genomes of other related species revealed that *O. gratissimum* is fully resolved in a clade with other *Ocimum* species classified to the family Lamiaceae.

### ARTICLE HISTORY

Received 11 December 2020  
Accepted 8 February 2021

### KEYWORDS

Complete chloroplast genome; Lamiaceae; *Ocimum gratissimum*; Tulsi

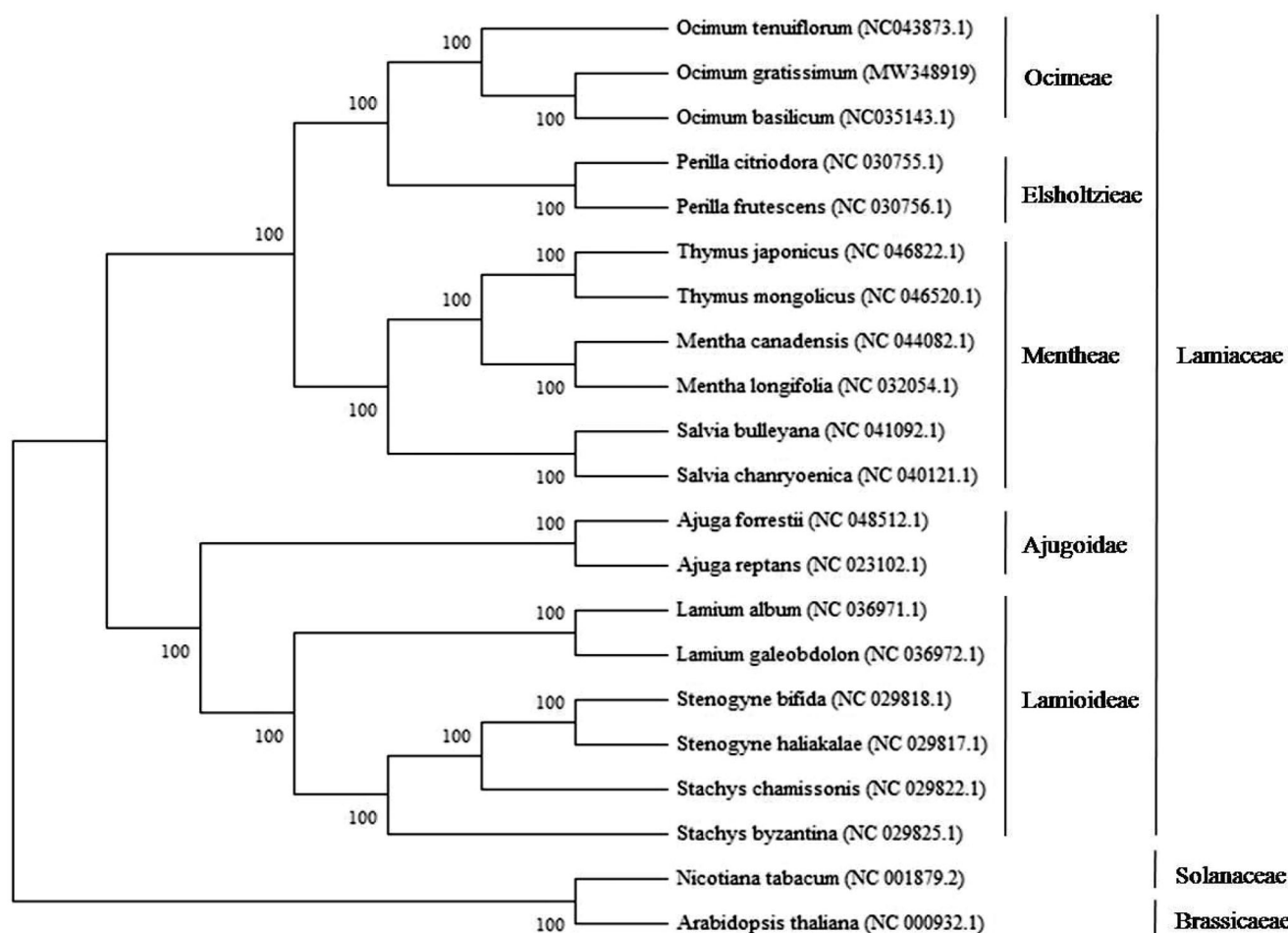
The genus *Ocimum* includes more than 150 species, and most of them are rich in therapeutic value (Willis 1966). The leaves of *Ocimum* species are known to have antimicrobial and anti-inflammatory properties, and to serve as a remedy for headache, fever, common cold, skin ailments, diabetes, vomiting, nausea, etc. Some species are also consumed as health food supplements due to their antioxidant and anti-allergic properties (Chopra et al. 1956; Prakash and Gupta 2005). *Ocimum gratissimum* L. is an important medicinal plant used to treat microbial infections, cephalalgia, nervous disorder, neuralgia, and rheumatism (Mohanraj et al. 2018). In this study, we sequenced and assembled the complete chloroplast genome of *O. gratissimum* using Illumina pair-end sequencing data to contribute to the systematics and evolutionary history of this taxon.

Fresh leaves of *O. gratissimum* were collected from the Foundation for Revitalization of Local Health Traditions (FRLHT), Bengaluru, India (GPS coordinates: 13°07'24.5"N, 77°32'52.3"E). The herbarium voucher specimen (SRMH000142) was deposited in the SRM IST Herbarium at the SRM Institute of Science and Technology, Kattankulathur, India (Herbarium curator: Dr. E. Pandian, Email: [pandiane@srmist.edu.in](mailto:pandiane@srmist.edu.in)). Total genomic DNA from *O. gratissimum* was extracted using the CTAB method (Doyle and Doyle 1987), with some modifications (Poovitha et al. 2016). A whole-genome DNA sequencing library was constructed using the Nextera XT Library Prep Kit. The library was sequenced on the Illumina NextSeq 500 platform (Illumina Inc., San Diego,

CA), and 3.63 Gb of paired-end sequencing data was obtained. The chloroplast genome of *O. gratissimum* was assembled using NovoPlasty (k-mer 31) with *O. tenuiflorum* L. (NC\_043873.1) as a reference seed sequence (Dierckxsens et al. 2017). The assembled *O. gratissimum* chloroplast genome was annotated with DOGMA (Wyman et al. 2004) and GeSeq (Tillich et al. 2017) using the chloroplast genomes of *O. tenuiflorum* (NC\_043873.1) and *O. basilicum* L. (NC\_035143.1) as reference sequences. The predicted transfer RNAs (tRNAs) were confirmed by tRNAscan-SE 2.0 (Lowe and Chan 2016). The complete chloroplast genome of *O. gratissimum* with annotations was submitted to GenBank (Accession No. MW348919), and the raw reads were deposited in the GenBank Sequence Read Archive (Accession No. SRR13206921).

The complete chloroplast genome of *O. gratissimum* is 152,469 bp with a mean coverage of 888x. It has a typical quadripartite structure, including a large single copy (LSC) region of 83,614 bp, a small single copy (SSC) region of 17,607 bp, and a pair of inverted repeats (IRs) of 25,624 bp. The chloroplast genome contains 85 protein-coding genes, 45 tRNA genes, and eight rRNA genes. Among them, six genes (*atpF*, *rpoC1*, *ndhK*, *rpl2*, *ndhA*, and *ndhB*) are single-intron genes, and two genes (*ycf3* and *clpP*) contain two introns. The overall GC content is 37.8%, while LSC, SSC, and IR regions are 35.6%, 31.7%, and 43.2%, respectively.

A neighbor-joining tree with 1000 bootstrap replicates was performed using MEGA v7.0 (Kumar et al. 2016) from the



**Figure 1.** Neighbor-joining (NJ) tree based on the whole chloroplast genome sequences of 21 species including *Nicotiana tabacum* L. and *Arabidopsis thaliana* L. as outgroups. The complete chloroplast genome sequences were aligned using MAFFT online version (<https://mafft.cbrc.jp/alignment/server/>) and subjected to generating NJ phylogenetic tree by MEGA v7.0 (Kumar et al. 2016). The bootstrap support values (>50%) from 1000 replicates are indicated in the nodes.

alignments created by the MAFFT program (Kato and Standley 2013). *Nicotiana tabacum* L. (Solanaceae) and *Arabidopsis thaliana* L. (Brassicaceae) were designated as outgroups, and 18 published chloroplast genomes from the Lamiaceae were included as ingroup taxa. The phylogenetic analysis fully resolved *O. gratissimum* in a clade with the closely related *O. basilicum*, in a clade also containing *O. tenuiflorum* (Figure 1). *Ocimum* was sister to the genus *Perilla* (Figure 1). This complete chloroplast genome of *O. gratissimum* can be subsequently used for phylogenetic analysis, DNA barcoding, and chloroplast genetic engineering studies of genus *Ocimum* and Lamiaceae family.

## Acknowledgment

We would like to acknowledge the SRM Institute of Science and Technology for providing research and computation facilities. We thank Dr. K. Ravikumar and Dr. N. Dhatchanamoorthy, Foundation for Revitalisation of Local Health Traditions (FRLHT), The University of Trans-Disciplinary Health Sciences and Technology, Bengaluru-560064, Karnataka, India, for their help in the collection and authentication of plant material.

## Disclosure statement

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

## Funding

We would like to acknowledge the funding from the SRM-Department of Biotechnology Ministry of Science and Technology Partnership Platform for Contemporary Research Services and Skill Development in Advanced Life Sciences Technologies [Grant No. BT/PR12987/INF/22/205/2015].

## Data availability statement

The data that support the findings of this study are openly available in GenBank at <https://www.ncbi.nlm.nih.gov/genome> under the accession numbers [NC\_048512, NC\_023102, NC\_036971, NC\_036972, NC\_029818, NC\_029817, NC\_029825, NC\_029822, NC\_030755, NC\_030756, NC\_035143, NC\_043873, NC\_046822, NC\_046520, NC\_044082, NC\_027784, NC\_041092, NC\_040121, and NC\_001879]. The complete chloroplast genome assembled in this study was deposited in GenBank under the accession number MW348919. The NGS sequencing data files are available from the BioProject, SRA, and Bio-Sample ID under the accession numbers PRJNA682904, SRR13206921, and SAMN17013753, respectively.

## References

- Chopra RN, Nayar SL, Chopra IC. 1956. Glossary of Indian medicinal plants. New Delhi (India): Council of Scientific and Industrial Research.
- Dierckx N, Mardulyn P, Smits G. 2017. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):1–9.

- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 33(7): 1870–1874.
- Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44(W1): W54–W57.
- Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, Samel A. 2018. IMPPAT: a curated database of Indian medicinal plants. *Phytochem Therap Sci Rep.* 8(4329):1–17.
- Poovitha S, Stalin N, Balaji R, Parani M. 2016. Multi-locus DNA barcoding identifies *matK* as a suitable marker for species identification in *Hibiscus L.* *Genome.* 59(12):1150–1156.
- Prakash P, Gupta N. 2005. Therapeutic uses of *Ocimum sanctum* Linn (Tulsi) with a note on eugenol and its pharmacological actions: a short review. *Ind J Physio Pharm.* 49:125–131.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organellar genomes. *Nucleic Acids Res.* 45(W1):W6–W11.
- Willis JC. 1966. A dictionary of flowering plants and ferns (revised 7th edn.). London: Cambridge University Press.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 20(17): 3252–3255.