


Complete chloroplast genome sequence and phylogenetic analysis of *Clerodendrum japonicum* (Thunb.) Sweet (Ajugoideae, Lamiaceae)

Gang Chen^a, Li-guo Wang^b and Ying-hua Wang^a 

^aCollege of Life Sciences, Zhaoqing University, Zhaoqing, Guangdong, China; ^bLaboratory of Germplasm Resources and Molecular Identification of Traditional Chinese Medicine, School of Pharmaceutical Sciences, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China

ABSTRACT

Clerodendrum japonicum (Thunb.) Sweet belongs to the genus *Clerodendrum* in the family Lamiaceae. It is an important medicinal plant with great ornamental and economic value. We sequenced and analyzed the complete chloroplast genome of *C. japonicum* by Illumina sequencing in this study. The full length of the complete chloroplast genome is 152,215 bp, containing a pair of inverted repeat regions of 25,705 bp (IRa and IRb) separated by a large single-copy region (LSC) of 83,491 bp and a small single-copy region (SSC) of 17,314 bp. The *C. japonicum* chloroplast genome encodes 131 genes, comprising 85 protein-coding genes, 37 tRNA genes, 8 rRNA genes, and 1 pseudogene. This study will be useful for further study on population, phylogenetic, and molecular genetic studies of this medicinal plant.

ARTICLE HISTORY

Received 5 February 2021
Accepted 12 June 2021

KEYWORDS

Chloroplast genome;
phylogenetic analysis;
Clerodendrum japonicum
(Thunb.) Sweet



Clerodendrum japonicum (Thunb.) Sweet is one species of the genus *Clerodendrum* (Lamiaceae), which is distributed predominantly in Guangdong, Fujian and Guangxi provinces in China. This species grows as 1–4 m shrubs with 4-angled branchlets that are pubescent to subglabrous with nodes that are sometimes villous. It bears thyrses with reddish bracts and bractlets that are usually reddish. The calyx is red, pubescent and deeply 5-lobed with a sandy, glandular outer surface. The corolla is a white or red tube with a 1.5–2.2 cm diameter that is slightly larger than the calyx (Wu and Raven 1994). It has been used in traditional medicine to treat upper respiratory tract infections, diarrhea, skin diseases, pneumonia, cough, fever, and conjunctivitis (Jia and Li 2005). The *C. japonicum* specimen used for plastome sequencing was cultivated in the Biological Garden of Zhaoqing University (N23°6', E112°30', Zhaoqing, China). The voucher specimen (No. BGCLSZU002) was deposited in the Herbarium of Zhaoqing University (IH code: GDZQU).

Total DNA of *C. japonicum* was extracted from a combination of young leaves taken from several different plants that were clonally propagated using the Plant Genomic DNA Kit (Tiangen, Beijing, China). And then it was sonicated into 300 bp fragments using the Covaris M220 (Covaris, Woburn, MA, USA) to construct shotgun sequencing libraries according to the TruSeq™ DNA Sample Prep Kit for Illumina. Whole genome sequencing was executed using the Illumina NovaSeq platform (Illumina, USA) (Genepioneer Biotechnologies Co. Ltd, Nanjing, China). Pair-end Illumina raw reads were cleaned from adaptors and barcodes and then quality filtered using Trimmomatic (Bolger et al. 2014). Putative reads were mapped

to the *C. mandarinorum* plastome (MN814861.1) as reference and Bowtie2 v2.2.4 (Langmead and Salzberg 2012) was used to exclude reads of nuclear and mitochondrial origins. The chloroplast genome was reconstructed with SPAdes 3.10.1 through *de novo* assembly method (Bankevich et al. 2012), and chloroplast contigs were concatenated into larger contigs using Sequencher 5.3.2 (Gene Codes Inc., Ann Arbor, MI, USA). The assembled *C. japonicum* plastome had a high average coverage depths of 3,653X (ranged from 955 to 4664X). Annotation of the chloroplast genomes were generated by CpGAVAS (Liu et al. 2012) and a circular representation was drawn using the online tool OGDRAW (Lohse et al. 2007). The complete chloroplast genome sequence has been submitted to GenBank with the accession number of MW307827.

The length of chloroplast genome sequence of *C. japonicum* is 152,215 bp, including two inverted repeat regions (IRa and IRb, each 25,705 bp) separated by a LSC (83,491 bp) region and a SSC (17,314 bp) region. The GC content of the overall chloroplast genome, IR regions, LSC, and SSC are 38.05, 43.27, 36.12 and 31.90%, respectively. The chloroplast genome contains 131 genes in total, including 85 protein-coding genes, 37 tRNAs, 8 rRNAs, and 1 pseudogene.

Plastomes from 34 species from Lamiaceae and its sister families available on NCBI GenBank were aligned with that from *C. japonicum* using MAFFT v7.427 (Katoh et al. 2005). Gaps were removed using trimAl with '-nogaps' v 1.4 (Capella-Gutierrez et al. 2009). Maximum-likelihood (ML) method was used to construct the phylogenetic tree with MEGA v7.0 (Kumar et al. 2016). As shown in the phylogenetic tree (Figure 1), *C. japonicum* forms a clade with only

CONTACT Ying-hua Wang  1050760740@qq.com  College of Life Sciences, Zhaoqing University, Zhaoqing Avenue, Zhaoqing, 526061, 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-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

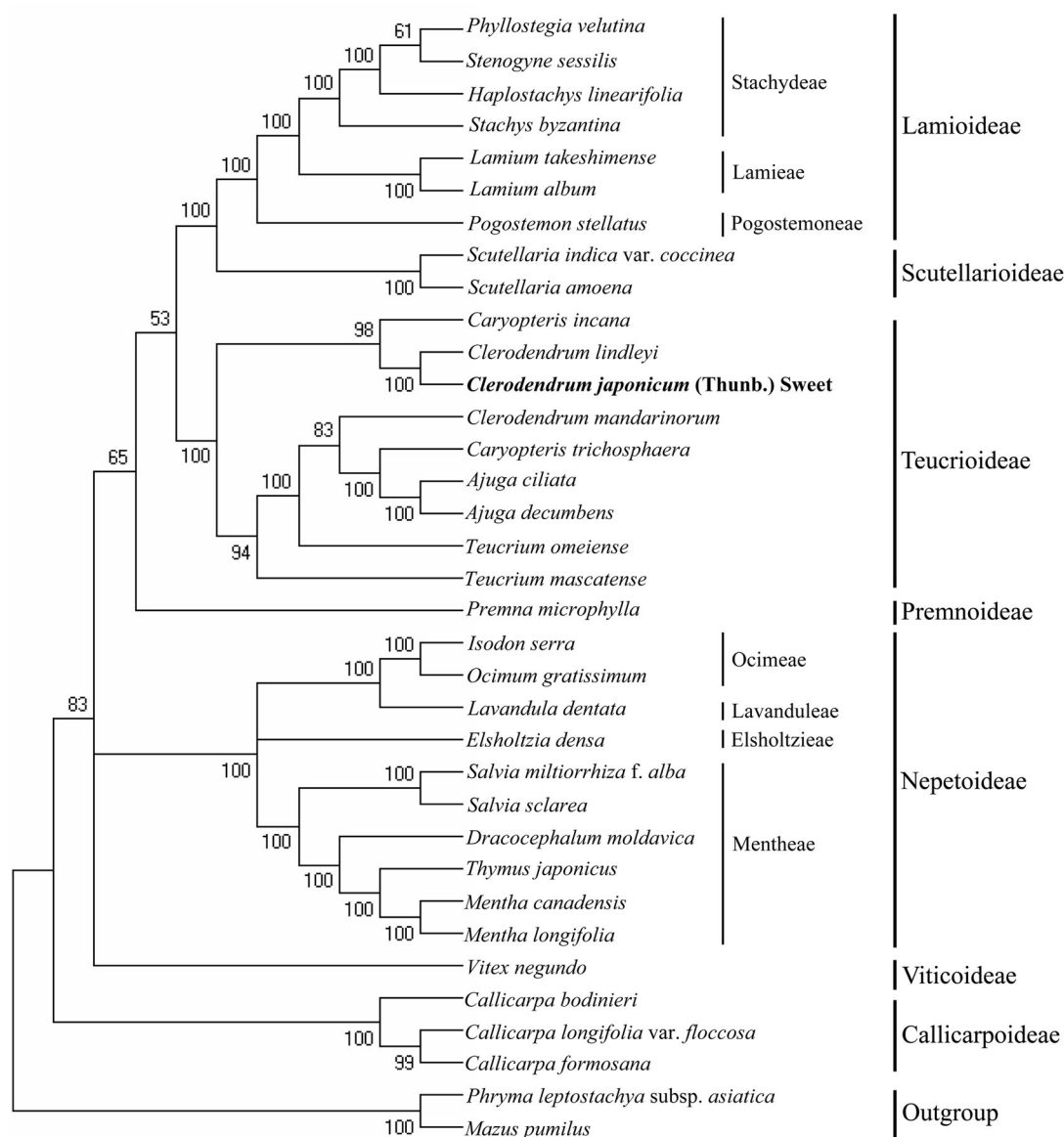


Figure 1. Phylogenetic reconstruction of 35 species based on whole chloroplast genome sequences. Bootstrap support values are indicated at each node. NCBI GenBank accession numbers are as follows *Ajuga ciliata* MN814853, *Ajuga decumbens* MN814854, *Callicarpa bodinieri* MW149077, *Callicarpa formosana* NC_052748, *Callicarpa longifolia* var. *floccosa* MW149076, *Caryopteris incana* MN918534, *Caryopteris trichosphaera* MN814860, *Clerodendrum japonicum* MW307827, *Clerodendrum lindleyi* MW411449, *Clerodendrum mandarinorum* MN814861, *Dracocephalum moldavica* MT457747, *Elsholtzia densa* MN793319, *Haplostachys linearifolia* KU724137, *Isodon serra* MT317099, *Lamium album* KY562589, *Lamium takeshimense* MN240520, *Lavandula dentata* MK791204, *Mazus pumilus* MT937187, *Mentha canadensis* MN047448, *Mentha longifolia* KU956042, *Ocimum gratissimum* MW348919, *Phryma leptostachya* subsp. *asiatica* MT948145, *Phyllostegia velutina* KU724134, *Pogostemon stellatus* KP718620, *Premna microphylla* KM981744, *Salvia miltiorrhiza* f. *alba* MT012420, *Salvia sclarea* MN520023, *Scutellaria amoena* MN128386, *Scutellaria indica* var. *coccinea* MN047312, *Stachys byzantina* KU724141, *Stenogyne sessilis* KU724135, *Teucrium mascatense* MH325132, *Teucrium omeiense* MN814871, *Thymus japonicus* MN867687, *Vitex negundo* MW366787.

Caryopteris incana and *Clerodendrum lindleyi*, and this clade is sister to another that contains *Clerodendrum mandarinorum*, *Caryopteris trichosphaera*, and the other Teucroioideae taxa for which whole chloroplast genomes have been sequenced (Li et al. 2016). The phylogenetic reconstruction suggests that genus *Clerodendrum* may not be monophyletic and/or there was a chloroplast capture event that resulted from an ancestral hybridization between early progenitors.

Author contributions

GC performed the experiments investigation, project administration, writing the original draft and data curation.

LW prepared the resources.

YW supervised the project and made revisions to the manuscript.

Disclosure statement

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

Funding

This study is funded by National Natural Science Foundation of China [81373901], National Survey on Chinese Materia Medica Resources of State Administration of Traditional Chinese Medicine [GZY-KJS-2018-284-001, GZY-KJS-2018-004, GZY-KJS-2019-004], and Special Subsidy Funds for National Public Health Service [CS (2016) No. 44].

ORCID

Ying-hua Wang  <http://orcid.org/0000-0002-5210-494X>

Data availability statement

The data that newly obtained at this study are available in the NCBI under accession number of MW307827 (<https://www.ncbi.nlm.nih.gov/nucleotide/MW307827>).

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–1973.
- Jia MR, Li XW. 2005. *Essentials of Chinese ethnic medicines*. Beijing (China): China Medical Science Press.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33(2): 511–518.
- 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.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Li B, Cantino PD, Olmstead RG, Bramley GL, Xiang CL, Ma ZH, Tan YH, Zhang DX. 2016. A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. *Sci Rep.* 6(1):34343–34318.
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. 2012. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom.* 13:715.
- Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet.* 52(5–6):267–274.
- Wu Z, Raven P. 1994. *Verbenaceae vol 17. Flora of China*. Beijing (China): Science Press.