

## Characterization of the complete chloroplast genome of *Salvia leucantha* (Lamiaceae)

Yan Zhou<sup>a</sup>, Hongrui Zhang<sup>a</sup>, Hongming Ping<sup>b</sup>, Yining Ding<sup>a</sup>, Saiwen Hu<sup>a</sup>, Guangyao Bi<sup>a</sup>, Cuicui Li<sup>a</sup>, Heming Li<sup>a</sup>, Yong Huang<sup>a</sup>, Linxiu Guo<sup>a</sup>, Xingli Ma<sup>a</sup> and Zhi Xia<sup>a</sup>

<sup>a</sup>College of Agronomy, Henan Agricultural University, Zhengzhou, China; <sup>b</sup>School of Computer Science, University of Nottingham Ningbo China, Ningbo, China

### ABSTRACT

*Salvia leucantha* (Lamiaceae) is an important horticultural plant with great ornamental and economic value. Here, we report the complete chloroplast genome of this species. The chloroplast genome was determined to be 151021 bp and the GC contents was 38.0%. The sequence includes a large single copy (LSC) region of 82,262 bp, a small single copy (SSC) region of 17,537 bp, and two separated inverted regions of 25,611 bp each. It contains 130 unique genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Based on the chloroplast genomes data of 26 species in *Salvia*, our result indicated that *S. leucantha*, *S. tiliifolia*, *S. hispanica*, and *S. splendens* formed one clade with Bootstrap = 100%. The four species belong to *Salvia* subgenus *Calosphace*, and *S. leucantha* was closely related to *Salvia tiliifolia* and *Salvia hispanica*. This result will facilitate population, genetic diversity and phylogenetic studies of *S. leucantha*.

### ARTICLE HISTORY

Received 19 July 2021  
Accepted 19 October 2021

### KEYWORDS

*Salvia leucantha*; *Salvia* subgenus *Calosphace*; chloroplast genome; Lamiaceae

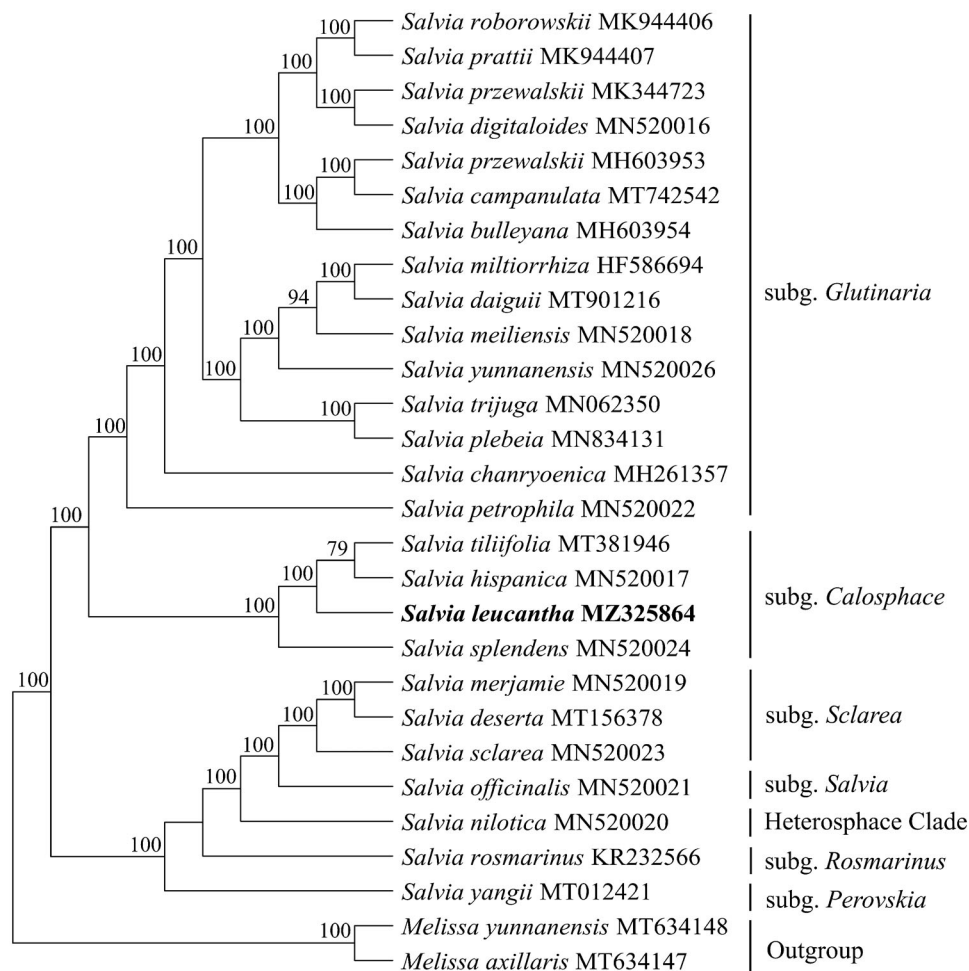
*Salvia leucantha* Cav. (common name Mexican bush sage) is a horticultural species belonging to the family Lamiaceae and originally distributed in Mexico (González-Gallegos et al. 2020). As one of introduced species, it is widely cultivated in China for the beauty of its long inflorescences. Neoclerodane diterpenoids and related compounds as major components have been isolated from the aerial parts of *S. leucantha* of the subgenus *Calosphace* (Rojas et al. 2010). In recent years, the complete chloroplast genome of some species in *Salvia* were reported (Zhang et al. 2020; Zhao et al. 2020). However, the complete chloroplast genome of *Salvia leucantha* was not reported. In this study, we first reported the complete chloroplast genome of *S. leucantha*, and inferred phylogenetic relationships of this species to other taxa of genus *Salvia* in Lamiaceae.

The fresh leaves of *S. leucantha* were collected in Zhongmou country, Henan Province, China (34°46'55"N, 114°0'20"E). A specimen was deposited at Herbarium of Henan Agricultural University (HEAC) (contact person and email: Zhi Xia, [xiazhiemail@126.com](mailto:xiazhiemail@126.com)) under the voucher number XZ-2021-01. The total genomic DNA was extracted following the CTAB method (Doyle and Doyle 1987). After DNA extraction, 1 µg genomic DNA was randomly fragmented by Covaris, followed by fragment selection by Agencourt AMPure XP-Medium kit to an average size of 200–400 bp. Then the genomic library (paired-end, PE 150 bp) was sequenced on BGISEQ-500 platform at Beijing Genomics Institute (Shenzhen, China). Totally, 2.12 Gb sequence reads

were obtained and used to assemble the chloroplast genome de novo by using NOVOPlasty (Dierckxsens et al. 2017) after trimming adaptor sequences and filtering low-quality reads. Then the assembled genome was annotated by using Geneious version 11.0.3 (Kearse et al. 2012) with manual adjustment. *Salvia hispanica* L. (GenBank accession number: MT083896) was used as reference plastid genome for assembling and annotation (Zhang et al. 2020). The annotated GenBank files of complete chloroplast sequence were submitted to GeneBank with the accession number MZ325864.

The complete chloroplast genome of *S. leucantha* has a length of 151,021 bp with a typical circular structure and a GC content of 38.0%. It consists of a large single copy region (LSC, 82,262 bp, 36.2% GC content), a small single-copy region (SSC, 17,537 bp, 31.7% GC content), and two inverted repeat regions (IR, 25,611 bp, 43.1% GC content). The genome harbors 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Among these, 15 genes (*trnK-UUU*, *rps16*, *trnG-UCC*, *atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *trnI-GUA*, *trnA-UGC*, and *ndhA*) had one intron, while three genes (*rps12*, *clpP*, and *ycf3*) contained two introns.

A total of 25 complete chloroplast genomes in *Salvia* (Lamiaceae) together with the chloroplast genomes of *S. leucantha* in this study were utilized to clarify the phylogenetic position of *S. leucantha*. Based on the result of Zhao et al. (2020), Li et al. (2016) and Drew et al. (2017), we selected the *Melissa officinalis* L. and *Melissa yunnanensis* C. Y. Wu et Y.



**Figure 1.** Maximum likelihood phylogenetic tree based on 26 complete chloroplast genome sequences of *Salvia*. The number on each node indicates the bootstrap value ■.

C. Huang as outgroup. The sequence alignment was conducted with MAFFT v7.3 (Kato and Standley 2013). A maximum likelihood analysis was performed with the RA × ML software (Stamatakis 2014) using 1000 bootstrap replicates. These analyses used the GTR substitution model with gamma-distributed rate heterogeneity among sites and the proportion of invariable sites estimated from the data. The phylogenetic tree indicates that all sampled species in *Salvia* formed a monophyletic group [Bootstrap (BS)=100%] (Figure 1). Our results are in line with the most recent phylogeny of the group (Fragoso-Martínez et al. 2018). Within the genus *Salvia*, *S. leucantha*, *Salvia tiliifolia* Vahl, *S. hispanica* and *Salvia splendens* Sellow ex Wied-Neuw formed a clade with Bootstrap = 100%. The four species were relatively closely related in Fragoso-Martínez et al. (2018) phylogeny and they all were included in the most recently derived clade: core Calosphace. It can be clarified that *S. leucantha* being close to *S. tiliifolia* and *S. hispanica*. The chloroplast genomes resource may be utilized for DNA barcoding, conservation genetics, and breeding of *S. leucantha* in the future.

## Disclosure statement

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

## Funding

This study was supported by the Scientific Research projects of Henan Provincial Education Department [Grant No. 22A360010].

## Data availability statement

The complete chloroplast 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. MZ325864. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA746149, SRR15115656, and SAMN20192381, respectively.

## References

- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Drew BT, González-Gallegos JG, Xiang CL, Kriebel R, Drummond CP, Walker JB, Sytsma KJ. 2017. *Salvia* united: the greatest good for the greatest number. *Taxon.* 66(1):133–114.
- Fragoso-Martínez I, Martínez-Gordillo M, Salazar GA, Sazatornil F, Jenks AA, García P, del R M, Barrera-Aveleida G, Benitez-Vieyra S, Magallón S, et al. 2018. Phylogeny of the neotropical sages (*Salvia* subg.

- Calosphace*; Lamiaceae) and insights into pollinator and area shifts. *Plant Syst Evol.* 304(1):43–55.
- González-Gallegos JG, Bedolla-García BY, Cornejo-Tenorio G, Fernández-Alonso JL, Fragoso-Martínez I, García-Peña M. d R, Harley RM, Klitgaard B, Martínez-Gordillo MJ, Wood JRI, et al. 2020. Richness and distribution of *Salvia* subgenus *Calosphace* (Lamiaceae). *Int J Plant Sci.* 181(8):831–856.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- 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.
- Li B, Cantino PD, Olmstead RG, Bramley GLC, 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: 34343.
- Rojas LB, Visba T, Morillo M, de Rojas YC, Arzola JC, Usubillaga A. 2010. The Volatile Constituents of *Salvia leucantha*. *Nat Prod Commun.* 5(6): 937–938.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9): 1312–1313.
- Zhang XJ, Chen C, Wang R, Yao Y, Liu LX, Zhang LY. 2020. Characterization of the complete chloroplast genome of *Salvia hispanica* (Lamiaceae). *Mitochondr DNA B.* 5(2):1748–1750.
- Zhao F, Drew BT, Chen YP, Hu GX, Li B, Xiang CL. 2020. The chloroplast genome of *Salvia*: genomic characterization and phylogenetic analysis. *Int J Plant Sci.* 181(8):812–830.