

Complete chloroplast genome of the oil-bearing shrub *Staphylea bumalda* DC (Staphyleaceae)

Han Zhao^a, Yue Qin^a, Gaopu Zhu^a, Tao Han^b and Panfeng Liu^a

^aResearch Institute of Non-Timber Forestry, Chinese Academy of Forestry, Zhengzhou, PR China; ^bState-Owned Tongbai Maoji Forest Farm Tongbai, PR China

ABSTRACT

Staphylea bumalda DC, belonging to family Staphyleaceae, is a woody understory tree that is both edible and medicinal and produces oil with high economic value. This study reports the first complete chloroplast genome sequence of *S. bumalda*. The complete chloroplast genome sequence of *S. bumalda* is 160,319 bp in length with an overall GC content of 32.79%, which is composed of a large single-copy region (LSC: 89,401 bp), a small single-copy region (SSC: 18,834 bp), and two inverted repeat regions (IR: 26,042 bp). A total of 130 genes were predicted in this genome, including 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. The phylogenetic analysis based on 14 complete chloroplast sequences from related species revealed that *S. bumalda* is a sister to *S. holocarpa*.

ARTICLE HISTORY

Received 14 February 2022
Accepted 30 December 2022

KEYWORDS

Staphylea bumalda;
chloroplast genome

Staphylea bumalda (Candolle 1825) is a shrub species in family Staphyleaceae mainly distributed throughout China, Japan, and Korea. Also called pearl flower, its floral buds contain a high content of crude protein, crude fat, and essential amino acids and are consumed as a vegetable in the mountainous regions of China for their unique flavor (Yang et al. 2013). The root of *S. bumalda* is used in Traditional Chinese Medicine for clearing heat and detoxification as well as reducing swelling, dispelling masses, moisturizing the lungs and cough suppression (Song 1999). Moreover, the seed oil yield is 32%, which is comparable to that of woody oil tree species such as hazelnut and walnut (Liu et al. 2008), which can be used to develop functional vegetable oils (Li et al. 2007). However, there are still few studies on the phylogenetic relationships among *S. bumalda* and related taxa. Here, we characterized the complete chloroplast genome sequence of *S. bumalda* using high throughput sequencing technology to provide a resource for exploring evolutionary relationships in Staphylea.

The fresh leaves of *S. bumalda* were collected from the State-owned Tongbai Maoji Forest Farm (32°23'45"N, 113°41'32"E) of Henan Province, China. The voucher specimen is deposited in the National Center for Forestry and Grassland Genetic Resources (contact person: Han Zhao, Hanzhao@caf.ac.cn; specimen code 1111C0003307005096). Total genomic DNA was extracted according to a modified CTAB protocol (Doyle and Doyle 1987). The total DNA quantity and purity were analyzed using Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, San Diego, CA, USA). The library was constructed with an insert length of 300 bp (±50 bp) and the genome sequencing was performed by Illumina HiSeq 2500 Platform (San Diego, CA, USA). After

removing low-quality reads and trimming adapter sequences, approximately 1.2 GB of high-quality clean reads remained. The filtered reads were assembled by SOAPdenovo (version: 2.04, <http://soap.genomics.org.cn/soapdenovo.html>) short sequence assembly software (Luo et al. 2015). The inner hole of the assembly result were repaired by GapCloser (version: 1.12, <http://soap.genomics.org.cn/soapdenovo.html>) software. The genome was annotated using the CpGAVAS2 pipeline (Shi et al. 2019), and adjusted using Geneious 20.2.2 (<http://www.geneious.com>). The new annotated chloroplast genome sequence was deposited in GenBank (Accession No. CRA005805).

The chloroplast genome sequence of *S. bumalda* was 160,319 bp in length, consisting of a small single-copy region (SSC: 18,834 bp), a large single-copy region (LSC: 89,401 bp), and two inverted repeat regions (IR: 26,042 bp), with the overall GC content of 32.79%. The chloroplast genome encoded a total of 130 genes, including 85 protein-coding genes, 37 tRNA genes, and eight rRNA genes. Twenty-seven introns were distributed across twenty-one genes contained. The gene content of the *S. bumalda* cp genome was nearly identical to those of related taxa. However, two genes, *LhbA* and *InfA* were lost in *S. bumalda* compared with other species.

To explore the phylogenetic relationships within tree species of related families and genera, a maximum likelihood phylogeny analysis was performed using IQ-TREE 2.1.3 with 1000 bootstrap replicates (Nguyen et al. 2015; Minh et al. 2020). A total of 14 species were used, including seven Staphyleaceae species, five *Sambucus* species, one *Bischofia* species, and one *Acer lucidum* species as outgroup. Genome

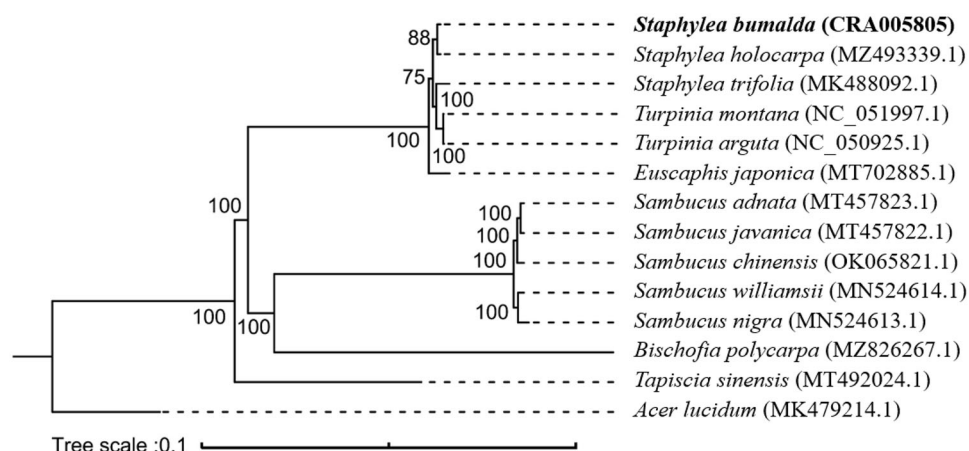


Figure 1. The maximum likelihood reconstruction of *Staphylea bumalda* and related species showing a monophyletic Staphyleaceae separate from taxa once considered related to *Staphylea*. Bootstrap values ($N = 1000$) are indicated at each node.

sequences were downloaded from the GenBank database and were aligned using MAFFT v7.0 (Kato and Standley 2013). As shown in the phylogenetic tree (Figure 1), the fourteen species were organized into five clusters. The species of Staphyleaceae and other families each formed separate clades, including those historically considered related to *Staphylea*. Staphyleaceae species have a relatively high degree of genetic similarity, and *S. bumalda* was sister to *S. holocarpa*. The topology shows *Tapiscia sinensis* is distantly related from *S. bumalda* and other *Staphylea*. This supports previous research describing the unique characters of *Tapiscia* that warrant its treatment as a separate family (Ding and Yu 1992; Shida and Tokuoka 2017).

Ethics statement

The sample collection for the study were permitted in 2017 by the State-owned Tongbai Maoji Forest Farm and the operations of acquiring these samples by the authors are in accordance with the management regulations.

Authors' contributions

Panfeng liu conceptualized the research program, designed experiments and coordinated the project. Yue Qin performed gene identification, annotation and analysis, and wrote the original manuscript; Gaopu Zhu and Han Zhao performed evolutionary analyses and critically revised the manuscript. Tao Han participated in sample survey and collection. All authors agree to be accountable for all aspects of the work.

Disclosure statement

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

Funding

This work was supported by National Forest Trees Germplasm Resources Platform [2003DKA21003]; National Center Forestry and Grassland Genetic Resources.

Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of CNCB at [https://ngdc.cnpc.ac.cn] (https://ngdc.cnpc.ac.cn) under the accession no. PRJCA007887. The associated BioProject, SRA, and Bio-Sample numbers are PRJCA007887, CRA005804, and SAMC545073 respectively.

References

- Ding SY, Yu ZY. 1992. The Taxonomic Significance of the Leaf Anatomical Structure of Staphyleaceae. *Plant Research*. 12(2):177–184.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 19:11–15.
- Kato K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
- Li X, Yuan FL, Yuan LP. 2007. Progressive research on the effect of polyunsaturated fatty acids in regulation of plasma lipids. *Anhui Medical and Pharmaceutical Journal*. 11(10):867–869.
- Liu ZX, Zhang HX, Liu T. 2008. Analysis of Oil Seeds and Evaluation of Functional Properties of Shenggu Oil Seeds. *Forestry Science*. 44(02):48–54.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2015. Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaSci*. 4:30.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 37(5):1530–1534.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, Liu C. 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res*. 47(W1):W65–W73.
- Shida T, Tokuoka T. 2017. Embryology and its character evolution in Staphyleaceae. *Plant Systematics and Evolution*. 303(10):1–13.
- Song LR. 1999. Chinese herb. Vol. 5. Shanghai: Shanghai Science and Technology Press; p. 217–218.
- Yang Y, Wang K, Gao RJ. 2013. Composition analysis and nutritional evaluation of 8 wild edible plants in Hanzhong. *J Shaanxi Normal Univ*. 3:6.