DATA NOTE
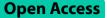
Assembly and characterization of the complete mitogenome of *Bauhinia purpurea* (Leguminosae)

Siqi Xie[1], Yong Chen[2], Danjing Zheng[2], Shukai Luo[2], Shiyuan Meng[2] and Yan Zhong[3*]

Abstract

Objective Mitochondrial genome sequences are very useful for understanding the mitogenome evolution itself and reconstructing phylogeny of different plant lineages. *Bauhinia purpurea*, a species from the legume family Leguminosae, is widely distributed in South China and has high ornamental value. Here, we sequenced and assembled the mitogenome of *B. purpurea* to provide a useful genetic resource for further evolutionary studies.

Data description We assembled and characterized the complete mitogenome of *B. purpurea* based on Illumina sequence data. The mitogenome size was 525,727 bp, and its GC content was 45.38%. A total of 35 protein-coding genes, 16 tRNA genes, and 3 rRNA genes were identified in the mitogenome. We also identified 124 pairs of repeats and 6 mitogenome sequences of plastid origin (MTPTs). These MTPTs range from 108 bp to 751 bp, covering 0.65% of the mitogenome.

Keywords Mitogenome, *Bauhinia purpurea*, Phylogenetics

Objective

The legume family (Leguminosae) ranks the third place in species richness across angiosperms, with about 20,000 species [1]. Many important crops or landscape trees are members of this family. *Bauhinia* is the largest genus in Cercidoideae, one of six subfamilies of Leguminosae [1]. Three species, *B. purpurea*, *B. variegata* and their hybrid *B. × blakeana*, in this genus have high horticultural values and are widely cultivated in South China [2, 3].

Although mitogenomes of more than 400 seed plants have been published, most focus on one species or distantly related species in a family, which hinders the study of evolution of mitogenome size and sequence, especially intergenic regions that show rapid divergence between closely related species. Comparing mitogenomes of closely related species within a genus may provide useful information on mitogenome size and sequence evolution [4–6]. Recently, the mitogenome of *B. variegata* has been published [7]. In this study, we assembled and characterized the complete mitogenome of its congeneric species, *B. purpurea*, based on Illumina sequencing data. Then, we made a preliminary mitogenome collinear analysis between the two species. The *B. purpurea* mitogenome sequence should be a useful resource for comparative mitogenome study in *Bauhinia*.

*Correspondence:
Yan Zhong
zhongy328@mail.sysu.edu.cn
[1] School of Life Sciences, State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Stress Biology, Sun Yat-sen University, Guangzhou 510275, China
[2] Guangzhou Institute of Forestry and Landscape Architecture, Guangzhou Collaborative Innovation Center on Science-Tech of Ecology and Landscape, Guangzhou 510405, China
[3] School of Ecology, Sun Yat-sen University, Shenzhen 518000, China

Xie *et al. BMC Genomic Data*   (2025) 26:6

Page 2 of 3

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data set 1	Illumina PE-150 reads of *Bauhinia purpurea*	SRA file (.*sra*)	NCBI Sequence Read Archive (http://identifiers.org/ncbi/insdc.sra:SRR29731951) [18]
Data file 1	The assembly and annotation of *Bauhinia purpurea* mitogenome.	Fasta file (.*fasta*)	https://identifiers.org/ncbi/insdc:PQ043320.1 [19]
Data file 2	Gene map of *Bauhinia purpurea* mitogenome.	Figure (.fig)	https://doi.org/10.6084/m9.figshare.26198948 [20]
Data file 3	Depth along the mitogenome of *Bauhinia purpurea*.	Figure (.fig)	https://doi.org/10.6084/m9.figshare.26199014 [21]
Data file 4	Collinear analysis between and within mitogenomes of *Bauhinia variegata* and *B. purpurea*.	Figure (.fig)	https://doi.org/10.6084/m9.figshare.26199059 [22]
Data file 5	Mitochondrial sequences of plastid origin (MTPTs) in the mitogenome of *Bauhinia purpurea*.	Figure (.fig)	https://doi.org/10.6084/m9.figshare.26200112 [23]
Data file 6	Maximum likelihood tree of 14 representative species in Leguminosae based on 23 concentrated mitochondrial genes.	Figure (.fig)	https://doi.org/10.6084/m9.figshare.26333164 [24]

Data description

The fresh young leaves of an individual of *B. purpurea* were collected from Sun Yat-sen University campus, Guangzhou, China. Fresh leaves were used for DNA extraction with a HiPure Plant DNA Mini Kit (Magen, Guangzhou, China). A genome library with an insert size of 350 bp was constructed and then sequenced on Illumina NovaSeq platform with paired-end reads of 150 bp. A total of 8 Gb raw data were generated. Clean data were obtained using Trimmomatic v0.39 [8] with default parameters. GetOrganelle v1.6.4 [9] with default parameters was used to assemble the mitogenome of *B. purpurea*. The protein-coding genes were annotated using GeSeq [10]. Annotation of ribosomal RNA and transfer RNA genes was carried out using RNammer v1.2 [11] and tRNAscan-SE v2.0 [12] with the "organelle" mode, respectively. All annotated genes were manually adjusted and confirmed. Repeats were identified using a python script ROUSFinder.py [13] with the minimal repeat size set to 30. Gene map of the mitogenome was drawn by PMGmap [14]. To identify the mitochondrial sequences of plastid origin (MTPTs), we compared the chloroplast genome sequence (GenBank accession number: NC_061218.1) and the mitogenome of *B. purpurea* using BLASTN [15] with *evalue* set to 1e-5. To construct the phylogenetic tree for 14 species from the four subfamilies of Leguminosae, these sequences of 23 shared mitochondrial protein-coding genes were concatenated and then aligned using MAFFT v7.520 [16]. The maximum likelihood (ML) method was implemented to build a phylogenetic tree in RAxML v8.2.12 [17] with a GAMMAAUTO substitution model and 1000 bootstrap replications.

The assembled circular mitogenome of *B. purpurea* is 525,727 bp in length and 45.38% in GC content. A total of 54 genes were annotated in the genome, including 35 protein-coding genes, 16 tRNA genes, and 3 rRNA genes. The protein-coding genes have a total length of 37,281 bp and account for 7.09% of the whole mitogenome (Table 1, Data files 1–2). The sequencing depth was relatively even across the mitogenome, with an average depth of 969.4 × (Table 1, Data file 3). A total of 124 repeat pairs were identified in the *B. purpurea* mitochondrial genome, and the longest repeat pairs are unusually large (60,619 bp). There are many collinear blocks and rearrangements between the mitogenomes of *B. purpurea* and *B. variegata*. The longest homologous block is about 57,198 bp (Table 1, Data file 4).

There are six highly homologous fragments between the mitochondrial and chloroplast genomes of *B. purpurea*, ranging from 108 bp to 751 bp in size and totally accounting for 0.65% of the mitogenome. Four complete genes of chloroplast origin (two *trnD-GUC* copies and two *trnW-CCA* copies) were identified on these six fragments (Table 1, Data file 5). In the maximum likelihood tree based on the 23 concatenated mitochondrial genes, *B. purpurea* is sister to *B. variegata* with 100% bootstrap support (Table 1, Data file 6).

The unusual large repeat identified in the mitogenome of *B. purpurea* has not been found in its congeneric species *B. variegata*, suggesting large repeats can be generated rapidly. The datasets produced in this study, along with available mitogenomes from three other subfamilies of Leguminosae, will benefit for phylogenetic studies of Leguminosae.

Limitations

Plant mitogenomes evolved rapidly in structure but slowly in sequence [25]. In the rate of synonymous substitutions, the mitochondrial genome evolved at a rate less than one-sixth that of the nuclear genome [26].

Xie *et al. BMC Genomic Data*   (2025) 26:6

Page 3 of 3

Therefore, it is not suitable for studying variation within species.

Abbreviations
GC   Guanine-Cytosine
tRNA   Transfer ribonucleic acid
rRNA   Ribosomal ribonucleic acid
DNA   Deoxyribonucleic acid

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Legume Phylogeny Working Group. Legume phylogeny and classification in the 21st century: a new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon. 2017;66(1):44–77.
2. Lau CP, Ramsden L, Saunders RM. Hybrid origin of *Bauhinia Blakeana* (Leguminosae: Caesalpinioideae), inferred using morphological, reproductive, and molecular data. Am J Bot. 2005;92(3):525–33.
3. Mak CY, Cheung KS, Yip PY, Kwan HS. Molecular evidence for the hybrid origin of *Bauhinia Blakeana* (Caesalpinioideae). J Integr Plant Biol. 2008;50(1):111–8.
4. Lai C, Wang J, Kan S, Zhang S, Li P, Reeve WG, Wu Z, Zhang Y. Comparative analysis of mitochondrial genomes of *Broussonetia* spp. (Moraceae) reveals heterogeneity in structure, synteny, intercellular gene transfer, and RNA editing. Front Plant Sci. 2022;13:1052151.
5. Wang M, Yu W, Yang J, Hou Z, Li C, Niu Z, Zhang B, Xue Q, Liu W, Ding X. Mitochondrial genome comparison and phylogenetic analysis of *Dendrobium* (Orchidaceae) based on whole mitogenomes. BMC Plant Biol. 2023;23(1):586.
6. Zhou S, Zhi X, Yu R, Liu Y, Zhou R. Factors contributing to mitogenome size variation and a recurrent intracellular DNA transfer in *Melastoma*. BMC Genomics. 2023;24(1):370.
7. Sun C, Chen Y, Zheng D, Zhong Y, Luo S, Meng S, Qian L, Wei D, Liu Y, Dai S. The complete mitochondrial genome of *Bauhinia variegata* (Leguminosae). Mitochondrial DNA Part B. 2024;9(1):128–32.
8. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
9. Jin J-J, Yu W-B, Yang J-B, Song Y, DePamphilis CW, Yi T-S, Li D-Z. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21(1):1–31.
10. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
11. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.
12. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016;44(W1):W54–7.
13. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. G3 (Bethesda). 2019;9(2):549–59.
14. Zhang X, Chen H, Ni Y, Wu B, Li J, Burzyński A, Liu C. Plant mitochondrial genome map (PMGmap): a software tool for the comprehensive visualization of coding, noncoding and genome features of plant mitochondrial genomes. Mol Ecol Resour. 2024;24(5):e13952.
15. Chen Y, Ye W, Zhang Y, Xu Y. High speed BLASTN: an accelerated MegaBLAST search tool. Nucleic Acids Res. 2015;43(16):7762–8.
16. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–80.
17. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
18. Xie S. Illumina sequencing data of *Bauhinia purpurea*. 2024. http://identifiers.org/ncbi/insdc.sra:SRR29731951
19. Xie S. Data file 1 for the Bauhinia purpurea mitochondrial genome. GenBank NCBI. 2024. https://identifiers.org/ncbi/insdc:PQ043320.1.
20. Xie S. Data files 2 for the mitochondiral genome of the plant Bauhinia purpurea. Figshare. 2024. https://doi.org/10.6084/m9.figshare.26198948.
21. Xie S. Data files 3 for the mitochondiral genome of the plant *Bauhinia purpurea*. Figshare. 2024 https://doi.org/10.6084/m9.figshare.26199014.
22. Xie S. Data files 4 for the mitochondiral genome of the plant Bauhinia purpurea. Figshare. 2024. https://doi.org/10.6084/m9.figshare.26199059.
23. Xie S. Data files 5 for the mitochondiral genome of the plant Bauhinia purpurea. Figshare. 2024. https://doi.org/10.6084/m9.figshare.26200112.
24. Xie S. Data files 6 for the mitochondiral genome of the plant Bauhinia purpurea. Figshare. 2024. https://doi.org/10.6084/m9.figshare.26333164.
25. Palmer JD, Herbon LA. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol. 1988;28:87–97.
26. Wolfe KH, Li W-H, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA. 1987;84(24):9054–8.

Publisher's note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.