

RESEARCH NOTE



## The complete chloroplast genome of the threatened *Dipentodon sinicus* (Dipentodontaceae)

MING-TAI AN<sup>1,2</sup>, XING-YONG CUI<sup>1</sup>, JIA-XIN YANG<sup>3</sup> and GUO-XIONG HU<sup>2,3\*</sup>

<sup>1</sup>College of Forestry, Guizhou University, Guiyang 550025, Guizhou, People's Republic of China

<sup>2</sup>Research Center for Biodiversity and Nature Conservation, Guizhou University, Guiyang 550025, Guizhou, People's Republic of China

<sup>3</sup>College of Life Sciences, Guizhou University, Guiyang 550025, Guizhou, People's Republic of China

\*For correspondence. E-mail: hugxgz@163.com.

Received 20 June 2018; revised 27 August 2018; accepted 11 September 2018; published online 7 February 2019

**Abstract.** *Dipentodon* is a monotypic genus of Dipentodontaceae and the only species, *Dipentodon sinicus*, is scattered in southwest China as well as adjacent Myanmar, northeast India and northern Vietnam. This species was evaluated as vulnerable in 'China Species Red List'. Here, we assembled and characterized the complete chloroplast (cp) genome of *D. sinicus* using Illumina sequencing data for the first time. The complete cp genome was 158,795 bp in length, consisting of a pair of inverted repeats of 26,587 bp, a large single-copy region of 87,233 bp and a small single-copy region of 18,388 bp. The genome encoded 113 unique genes, including 79 protein-coding genes, 30 tRNA genes and four rRNA genes. Phylogenetic analysis based on 16 complete cp genome sequences indicated that *D. sinicus* is a member of Huerteales, consistent with its position in the latest classification of flowering plants (AGP IV).

**Keywords.** chloroplast genome; Dipentodontaceae; Huerteales; *Dipentodon sinicus*.

*Dipentodon* is a monotypic genus of Dipentodontaceae, with only *Dipentodon sinicus* included in it (Ma and Bartholomew 2008). This species is a semi-evergreen shrub (rarely small tree), usually seen in mountain slopes, riverbanks and trail sides. Currently, *D. sinicus* is sporadically distributed in southwest China as well as adjacent Myanmar, northeast India and northern Vietnam (Ma and Bartholomew 2008; Yuan *et al.* 2008). Due to its rapid population decline, *D. sinicus* has been evaluated as vulnerable (VU) in 'China Species Red List' (Wang and Xie 2004). Although *D. sinicus* is listed as lower risk/least concern in the latest IUCN Red List of Threatened Species (Version 2018-1, see [www.iucnredlist.org](http://www.iucnredlist.org)), the wild resources of the species are not effectively protected with cutting and habitat destruction in China. To date, no data are available regarding the chloroplast (cp) genome of *D. sinicus*. In this study, we assembled and characterized the complete cp genome of *D. sinicus* for the first time.

Leaf samples of *D. sinicus* were collected from Leigongshan National Nature Reserve, Guizhou, southwestern China and the specimens were deposited in the Guizhou

University Museum (Voucher: Hu *et al.* 0409). Total DNA was extracted from a single leaf using modified CTAB (cetyl trimethylammonium bromide) extraction protocol (Doyle and Doyle 1987). Approximately 5 µg DNA was prepared for library construction and sequenced on Illumina HiSeq 2500 platform (San Diego, USA). About 10 Gb pair-end (150 bp) raw short sequence data were obtained and Novoplasty (Dierckxsens *et al.* 2017) was used to extract and assemble cp genome. We set the recommended 39 as K-mer value and used the sequence of closely related species *Tapiscia sinensis* (NC036960) as a seed input in the analysis. To validate the assembly, PCR amplifications and Sanger sequencing were conducted to verify the four junctions between IRs and large single-copy region (LSC) / small single-copy region (SSC). The cp genome annotation was performed using DOGMA (Wyman *et al.* 2004), coupled with manual correction for protein-coding region (CDS) boundaries. The tRNA genes were identified using DOGMA and tRNAscan-SE 1.21 (Schattner *et al.* 2005). The circular genome map was drawn using OGDRAW program (Lohse *et al.* 2013).

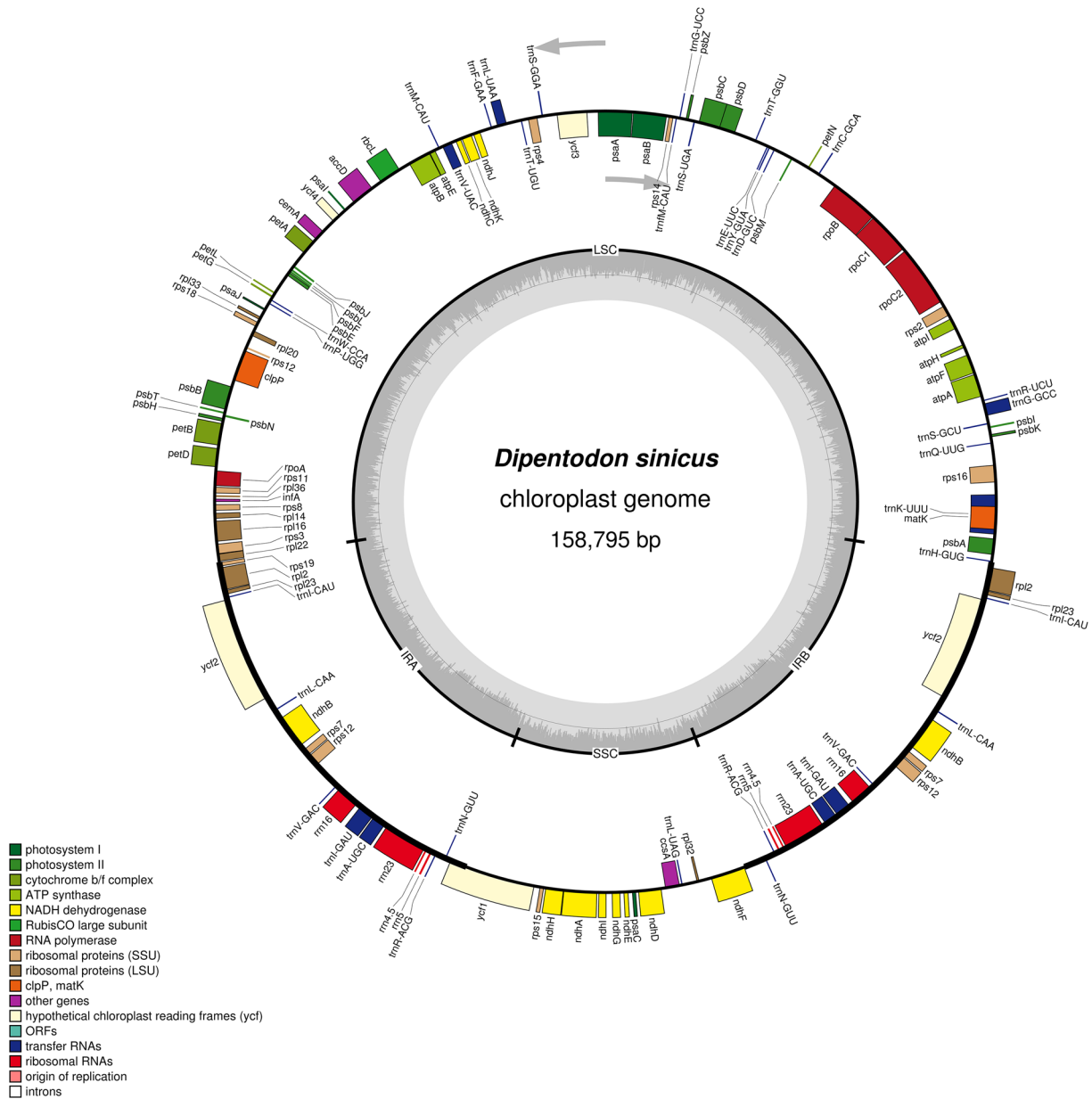


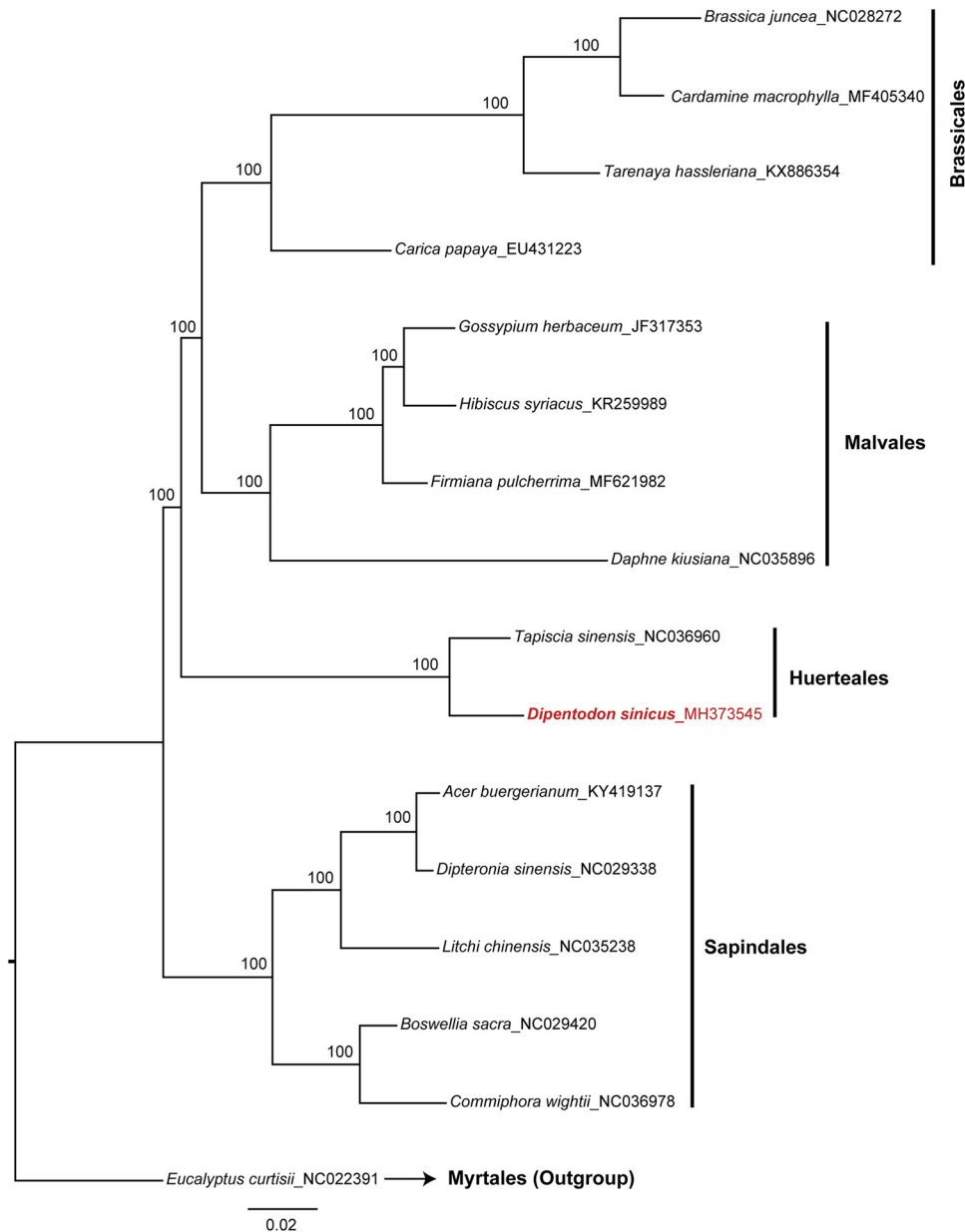
Figure 1. Gene map of the *D. sinicus* cp genome.

Table 1. Taxa used in phylogenomic analysis in this study.

Scientific name	Family	Order	GenBank accession no.
<i>Brassica juncea</i> (L.) Czern.	Brassicaceae	Brassicales	NC028272
<i>Cardamine macrophylla</i> Schur	Brassicaceae	Brassicales	MF405340
<i>Tarenaya hassleriana</i> (Chodat) Iltis	Cleomaceae	Brassicales	KX886354
<i>Carica papaya</i> L.	Caricaceae	Brassicales	EU431223
<i>Gossypium herbaceum</i> L.	Malvaceae	Malvales	JF317353
<i>Hibiscus syriacus</i> L.	Malvaceae	Malvales	KR259989
<i>Firmiana pulcherrima</i> H.H. Hsue	Malvaceae	Malvales	MF621982
<i>Daphne kiusiana</i> Miq.	Thymelaeaceae	Malvales	NC035896
<i>Tapiscia sinensis</i> Oliv.	Staphyleaceae	Huerteales	NC036960
<i>D. sinicus</i> Dunn	Dipentodontaceae	Huerteales	MH373545
<i>Acer buergerianum</i> Miq.	Sapindaceae	Sapindales	KY419137

**Table 1.** (Contd)

Scientific name	Family	Order	GenBank accession no.
<i>Dipteronia sinensis</i> Oliv.	Sapindaceae	Sapindales	NC029338
<i>Litchi chinensis</i> Sonn.	Sapindaceae	Sapindales	NC035238
<i>Boswellia sacra</i> Flueck.	Burseraceae	Sapindales	NC029420
<i>Commiphora wightii</i> (Arn.) Bhandari	Burseraceae	Sapindales	NC036978
<i>Eucalyptus curtisii</i> Blakely & C.T. White	Myrtaceae	Myrtales	NC022391

**Figure 2.** Maximum likelihood phylogenetic tree inferred from 16 complete cp genomes. The position of *D. sinicus* is highlighted in red and numbers above each node are bootstrap support values.

The complete cp genome of *D. sinicus* (GenBank accession number: MG797568) was 158,795 bp in length, with a biased GC content of 37.5%. The complete cp genome shows a typical quadripartite structure with a

pair of inverted repeats (IRs) of 26,587 bp, separated by a large single-copy region (LSC) of 87,233 bp and a small single-copy region (SSC) of 18,388 bp (figure 1). A total of 130 genes were annotated in the cp genome of

*D. sinicus*, of which 113 were unique, consisting of 79 protein-coding genes, 30 tRNA genes and four rRNA genes. Six protein-coding, seven tRNA and all four rRNA genes were duplicated in the IR regions. In total, 18 intron-containing genes were annotated in the cp genome of *D. sinicus* of which three genes (*clpP*, *rps12* and *ycf3*) include two introns and the rest include a single intron.

To reveal the phylogenetic position of *D. sinicus* within the Malvids lineage, we performed a phylogenomic analysis using RAxML-HPC2 on XSEDE v.8.1.11 (Stamatakis 2014) as implemented on the CIPRES computer cluster (<http://www.phylo.org/>) (Miller et al. 2010). Depending on the available data from GenBank, we selected one or three closely related families from five different orders within the Malvids, respectively. In total, the dataset includes 16 complete cp genomes representing 10 families (table 1). The GTRCAT model was used for analyses and bootstrapping. Except for setting bootstrap iterations (-#| -N) to 1000, other parameters followed default settings (Hu et al. 2018). Our result confirmed that the *D. sinicus* was a member of Huerteales (figure 2), consistent with its position in the latest classification of flowering plants (AGP IV) (Angiosperm Phylogeny Group 2016).

Complete cp genome has been increasingly used to solve phylogenetic relationships at different taxonomical levels within seed plants (Parks et al. 2009; Fu et al. 2017; Zhang et al. 2017; Geng et al. 2018). The complete cp genome of *D. sinicus* can be used for further phylogenomic studies of Huerteales and will provide fundamental information to effectively conserve important and threatened tree species.

#### Acknowledgements

This work was supported by the Natural Science Foundation of Guizhou Province (Qian-Ke-He [2016]1049 and [2016]1033), Application Fundamentals Major Special Projects (Qian-Ke-He JZ [2014] 2002) and the Second Investigation on National Key Protected Wild Plant Resources in Guizhou Province (Lin-Hu-Fa-No. [2012] 87 and Qian-Lin-Hu-Tong-No. [2013] 251).

#### References

Angiosperm Phylogeny Group. 2016 An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot. J. Linn. Soc.* **181**, 1–20.

- Dierckxens N., Mardulyn P. and Smits G. 2017 NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18.
- Doyle J. J. and Doyle J. L. 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15.
- Fu C. N., Li H. T., Milne R., Zhang T., Ma P. F., Yang J. et al. 2017 Comparative analyses of plastid genomes from fourteen *Cornales* species: Inferences for phylogenetic relationships and genome evolution. *BMC Genomics* **18**, 956.
- Geng Y. F., Hu G. X., Wang S. and Xu J. C. 2018 Complete chloroplast genome of the threatened *Rhoiptelea chiliantha* (Juglandaceae s.l.). *Conserv. Genet. Resour.* (in press).
- Hu G. X., Takano A., Drew B. T., Liu E. D., Soltis D. E., Soltis P. S. et al. 2018 Phylogeny and staminal evolution of *Salvia* (Lamiaceae, Nepetoideae) in East Asia. *Ann. Bot.* **122**, 649–668.
- Lohse M., Drechsel O., Kahlau S. and Bock R. 2013 OrganellarGenomeDRAW: A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl. Acids Res.* **41**(Web Server issue), W575–W581.
- Ma J. S. and Bartholomew B. 2008 Dipentodontaceae. In *Flora of China* (ed. C. Y. Wu, P. H. Raven and D. Y. Hong), p. 359. Science Press, Beijing & Missouri Botanical Garden Press, St Louis.
- Miller M. A., Pfeiffer W. and Schwartz T. 2010 *Creating the CIPRES science gateway for inference of large phylogenetic trees*. Gateway Computing Environments Workshop (GCE), 2010. IEEE (<http://www.phylo.org/>).
- Parks M., Cronn R. and Liston A. 2009 Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **7**, 84.
- Schattner P., Brooks A. N. and Lowe T. M. 2005 The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, 686–689.
- Stamatakis A. 2014 RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
- Wang S. and Xie Y. 2004 *China species red list*. Higher Education Press, Beijing.
- Wyman S. K., Jansen R. K. and Boore J. L. 2004 Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**, 3252–3255.
- Yuan Q. J., Zhang Z. Y., Peng. H. and Ge S. 2008 Chloroplast phylogeography of *Dipentodon* (Dipentodontaceae) in southwest china and northern Vietnam. *Mol. Ecol.* **17**, 1054–1065.
- Zhang M. Y., Fritsch P. W., Ma P. F., Wang H., Lu L. and Li D. Z. 2017 Plastid phylogenomics and adaptive evolution of *Gaultheria* series *Trichophyllae* (Ericaceae), a clade from sky islands of the Himalaya-Hengduan Mountains. *Mol. Phylogenet. Evol.* **110**, 7–18.

Corresponding editor: MANOJ PRASAD