

Characterization of the complete chloroplast genome of the endangered species *Carya sinensis* (Juglandaceae)

Yiheng Hu¹ · Xi Chen¹ · Xiaojia Feng¹ · Keith E. Woeste² · Peng Zhao¹

Received: 16 August 2016 / Accepted: 22 August 2016 / Published online: 1 September 2016
© Springer Science+Business Media Dordrecht 2016

Abstract *Carya sinensis* (Chinese Hickory, beaked walnut, or beaked hickory) is an endangered species that needs urgent conservation action. Here, we reported the complete chloroplast (cp) genome sequence and the genomic features of the *C. sinensis* cp, which is the first complete cp genome of any member of *Carya*. The *C. sinensis* cp genome was 160,195 bp, containing a large single copy region (LSC) of 89,541 bp and a small single copy region (SSC) of 18,538 bp separated by a pair of 26,058 bp inverted repeat regions (IRs). The genome contained 135 genes, including 87 protein-coding genes (79 species), 40tRNA genes (30tRNA species) and eight ribosomal RNA genes (four rRNA species). Most genes occurred as a single copy; 22 were duplicated. The overall GC content of the *C. sinensis* cp genome was 36.3%; the corresponding GC content

values of the LSC, SSC and IR regions were 33.9, 30.0 and 42.5%, respectively. A maximum likelihood (ML) phylogenetic analysis showed that *C. sinensis* is closely related to *Juglans*, as expected based on taxonomic affinity. The complete cp genome of *C. sinensis* will be useful for further investigations of genetic diversity and conservation of this endemic relict woody plant.

Keywords Chinese hickory · Annamocarya · Conservation · Phylogeny · Endemic · Red list

Chinese Hickory (*Carya sinensis* Dode) is a rare tertiary relict tree species belonging to the Juglandaceae family. This species has been regarded as an endangered plant because it occurs in small populations scattered along river-banks as a component of sub-tropical broad-leaf forests of southern China and northern Vietnam (Zhang et al. 2013a). Numbers of *C. sinensis* have dramatically decreased and the species needs urgent conservation. It was included in The IUCN Red List of Threatened Species (<http://www.iucnredlist.org/details/32321/0>). In China the species has been difficult to locate at recorded localities (<http://www.iucnredlist.org/details/32321/0>). Occurring in small groves or as isolated individuals, this relict species is recorded from a few forest localities in north Guizhou, north-west Guangxi, south-east Guizhou, south-east Yunnan in China, and from a few localized populations in north and west Viet Nam (IUCN Red List of Threatened Species, Version 2015-4; Sun 1998).

Chloroplast genomes are widely used in phylogeny (Jansen et al. 2007), DNA barcoding (Stanford et al. 2000), species conservation and studies of genome evolution (Moore et al. 2007). Here, we reported the complete chloroplast genome sequence of the *C. sinensis* based on next-generation sequencing. The annotated cp genome has been deposited

✉ Peng Zhao
pengzhao@nwu.edu.cn
Yiheng Hu
berlin@stumail.nwu.edu.cn
Xi Chen
781291157@qq.com
Xiaojia Feng
332146704@qq.com
Keith E. Woeste
kwoeste@fs.fed.us

¹ Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, College of Life Sciences, Northwest University, Taibaibei Road 229, Xi'an 710069, Shaanxi Province, China

² USDA Forest Service Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, 715 West State Street, West Lafayette, IN 47907, USA

into GenBank with the accession number KX671976, which is the first complete cp genome of any member of *Carya*.

The fresh leaves of a single individual of *C. sinensis* were sampled from Sandu (Guizhou, China; 107.811°E, 25.944°N), and used for extraction of total genomic DNA using a modified CTAB method (Zhao and Woeste 2011). DNA samples and a voucher specimen of *C. sinensis* were deposited in the Evolutionary Botany Laboratory, College of Life Sciences, Northwest University (Xi'an, Shaanxi, China). Whole-genome sequencing was performed with

150 bp pair-end reads on the Illumina Hiseq Platform (Illumina, San Diego, CA). In total, about 13 million (13,878,699) high quality clean reads were obtained and used for the cp genome assembly using SPAdes version 3.9.0 (Bankevich et al. 2012). The resulting scaffolds were linked based on overlapping regions after being aligned to *Juglans regia* (GenBank accession number KT963008) (Hu et al. 2016) and visualized in Geneious version 8.0.2 (Kearse et al. 2012). Annotation was performed with Dual Organelar Genome Annotator (DOGMA) software (Wyman et al.

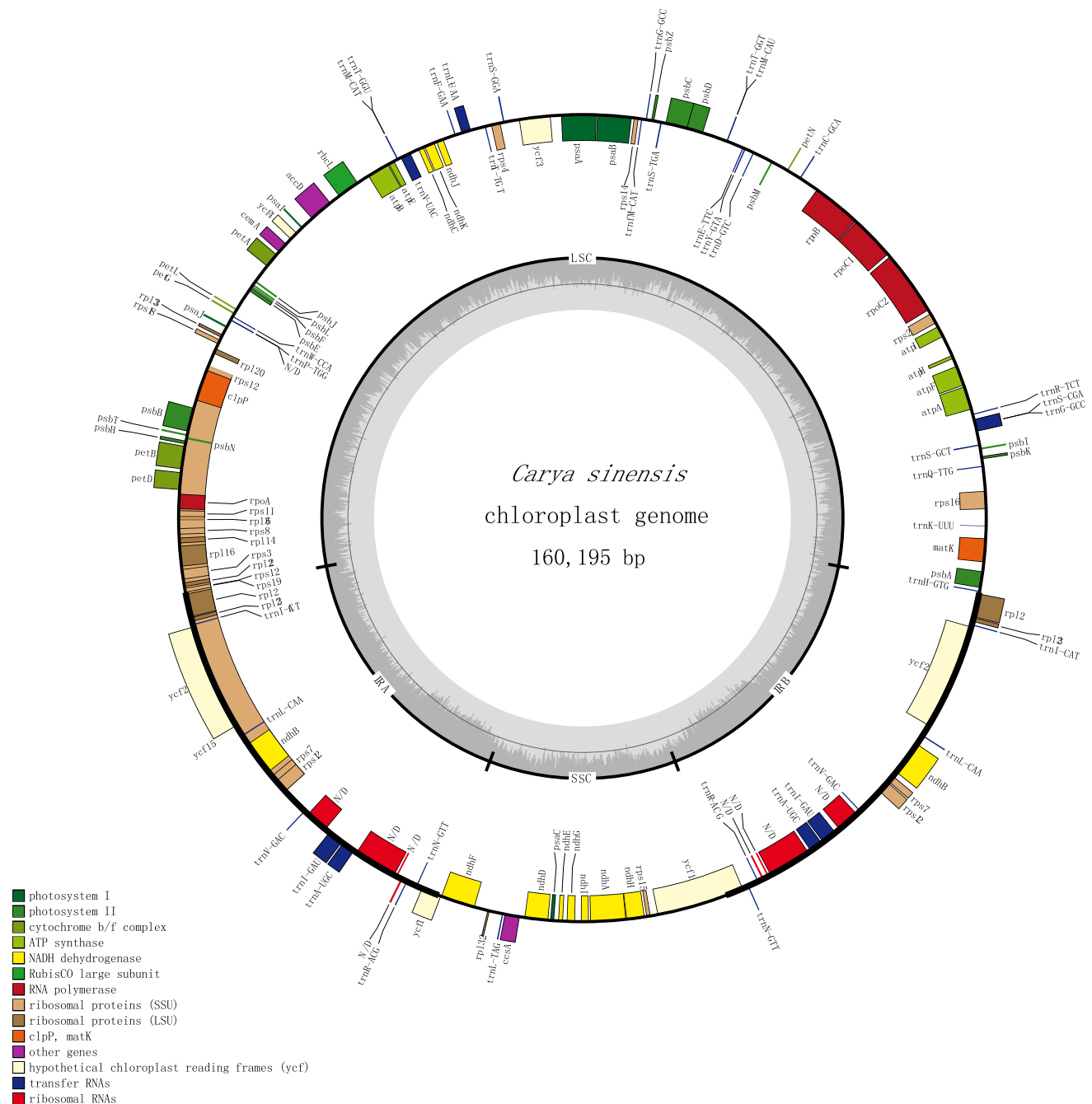


Fig. 1 Gene map of the *C. sinensis* chloroplast genome

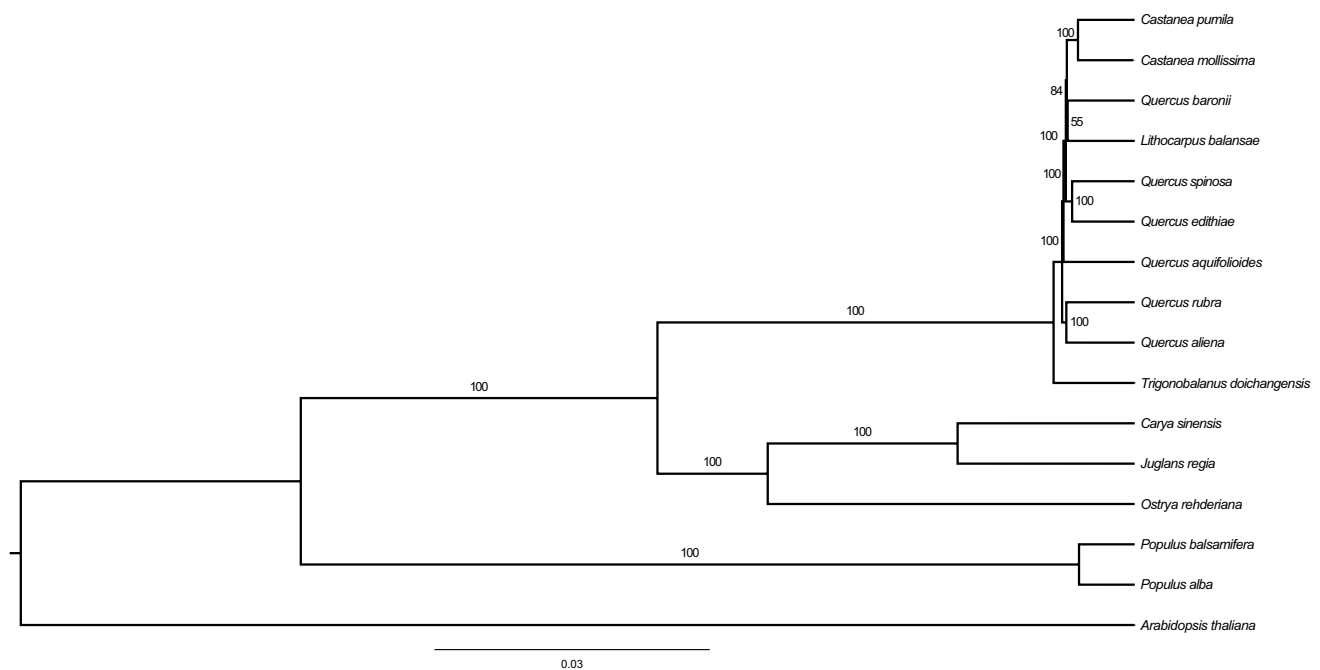


Fig. 2 Phylogeny of 16 species based on the maximum likelihood (ML) analysis of their whole chloroplast genomes. The 16 species can be divided into four independent clades: Fagales, Salicales and Brassicales. *Arabidopsis thaliana* (Brassicales) was used as an outgroup. Bootstrap support values are given at the nodes. Chloroplast genome accessions used: *Arabidopsis thaliana*: NC000932; *Populus alba*: AP008956; *Populus balsamifera*: NC024735; *Lithocarpus*

balansae: KP299291; *Quercus aliena*: NC026790; *Quercus aquifolioides*: NC026913; *Quercus baronii*: KT963087; *Quercus edithiae*: KU382355; *Quercus rubra*: JX970937; *Quercus spinosa*: NC026907; *Trigonobalanus doichangensis*: KF990556; *Juglans regia*: KT963008; *Castanea mollissima*: NC014674; *Castanopsis echinocarpa*: NC023801; *Ostrya rehderiana*: NC028349

2004). We corrected the annotation with Geneious (Kearse et al. 2012). A map of the genome was generated using OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/>) (Lohse et al. 2013). A maximum likelihood (ML) tree with 1000 bootstrap replicates was inferred using RaxML version 8.0.0 (Stamatakis 2014) from alignments created by the MAFFT version 7.304 (Katoh and Standley 2013) using chloroplast genomes of 16 species.

The complete assembled chloroplast genome of *C. sinensis* is typical in its general structure with a pair of IRs (26,058 bp), a LSC region (89,541 bp), and a SSC region (18,538 bp) (Fig. 1). The GC content of the chloroplast DNA was 36.3%, similar to the other reported chloroplast genomes from the family Juglandaceae (Hu et al. 2016). The GC content of the LSC and SSC were 33.9 and 30.2% respectively. The chloroplast genome of *C. sinensis* encodes 113 genes, including 79 protein coding genes, 30 transfer RNAs (tRNA), and 4 ribosomal RNAs (rRNA). Eighteen genes are duplicated in the IR. Fifteen genes contained one intron (*rps16*, *atpF*, *rpoC1*, *petB*, *petD*, *rpl16*, *rpl2*, *ndhB*, *ndhA*, *trnS-GCU*, *trnL-UAA*, *trnV-UAC*, *trnI-GAU*, *trnA-UGC*), while *ycf3* and *clpP* each contains two introns. The 5'-end exon of the *rps12* gene is located in the LSC region while its intron and 3'-end exon are situated in the IR region.

The phylogenetic tree constructed from genomes of 16 species (Fig. 2) showed that *C. sinensis* (Juglandaceae) was most closely related to *Juglans* and nearer to *Ostrya* (Betulaceae) than *Quercus* (Fagaceae) as expected based on previous taxonomy (Lu et al. 1999; Zhang et al. 2013b; <http://www.mobot.org/mobot/research/apweb/orders/fagalesweb.htm>; http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=200006095). The chloroplast genome of *C. sinensis* will provide useful resources for further study of the species' genetic diversity, its relationship to other *Carya*, and the conservation of this highly endangered species.

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 41471038, No. 31200500 and No. J1210063), the Program for Excellent Young Academic Backbones funding by Northwest University, Changjiang Scholars and Innovative Research Team in University (IRT1174). Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable.

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Pyshkin AV (2012) SPAdes: a new genome assembly

- algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–477
- Hu Y, Woeste KE., Dang M, Zhou T, Feng X, Zhao G, Zhao P (2016) The complete chloroplast genome of common walnut (*Juglans regia*). *Mitochondrial DNA Part B* 1–2
- Jansen RK, Cai Z, Raubeson LA, Daniell H, Leebens-Mack J, Müller KF, Lee SB (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc Natl Acad Sci USA* 104(49):19369–19374
- 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, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) Organellar Genome DRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:W575–W581
- Lu AM, Stone DE, Grauke LJ (1999) Juglandaceae. In Wu ZY, Raven PH (eds) *Flora of China*, volume 4, *Cycadaceae* through *Fagaceae*, Science Press, Beijing, pp 277–285
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proc Natl Acad Sci USA* 104(49):19363–19368
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Stanford AM, Harden R, Parks CR (2000) Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *Am J Bot* 87(6):872–882
- Sun W (1998) *Carya sinensis*. The IUCN Red List of Threatened Species 1998: e.T32321A9697238
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255
- Zhang JB, Li RQ, Xiang XG., Manchester SR, Lin L, Wang W, Chen ZD (2013a) Integrated fossil and molecular data reveal the biogeographic diversification of the eastern Asian-eastern North American disjunct hickory genus (*Carya Nutt.*). *PLoS One* 8(7):e70449
- Zhang ZY, Pang XM, Han JW, Wang Y, Li YY (2013b) Conservation genetics of *Annamocarya sinensis* (Dode) Leroy, an endangered endemic species. *Genet Mol Res* 12(3):3965–3974
- Zhao P, Woeste KE (2011) DNA markers identify hybrids between butternut (*Juglans cinerea* L.) and Japanese walnut (*Juglans ailantifolia* Carr.). *Tree Genet Genomes* 7(3):511–533