

Characterization of the complete chloroplast genome of wheel wingnut (*Cyclocarya paliurus*), an endemic in China

Yiheng Hu¹ · Jing Yan¹ · Xiaojia Feng¹ · Meng Dang¹ · Keith E. Woeste² · Peng Zhao¹

Received: 26 November 2016 / Accepted: 27 December 2016 / Published online: 12 January 2017
© Springer Science+Business Media Dordrecht 2017

Abstract The wheel wingnut (*Cyclocarya paliurus*) is an endemic species distributed in eastern and central China. *Cyclocarya* is a woody genus in the Juglandaceae used in medicine and horticulture. The complete chloroplast genome of *C. paliurus* was sequenced using the Illumina Hiseq 2500 platform. The total genome length was 160,562 bp comprised of a large single copy (LSC) region of 90,007 bp, an inverted region A (IRa) of 26,039 bp, a short single copy (SSC) region of 18,477 bp, and an inverted region B (IRb) region of 26,039 bp. The genome consisted of 137 annotation genes, with 80 unique coding regions and nine coding regions which were repeated in the inverted region. Among these, 40 unique tRNAs with nine duplications, and 4 rRNAs were all duplicated in the IR

region. An analysis of chloroplast phylogeny strongly supported the placement of *C. paliurus* near to *Juglans regia*.

Keywords Wheel wingnut · Illumina sequencing · Complete chloroplast genome · Phylogenetic

Cyclocarya paliurus (Batal.) Iljinskaja, known commonly as wheel wingnut, is the sole member of the genus *Cyclocarya* Iljinskaja (Juglandaceae). It is a relict endemic tree in subtropical China (mainly 24°N to 40°N and from about 105°E to 122°E) (Crane and DuVal 2013). Wheel wingnut produces fruits comprised of a nutlet surrounded by a circular wing; the tree itself is used in herbal medicine and the landscape horticulture trade. However, due to growth difficulties and seed dormancy time is long (2 years) in natural environment, germplasm resource of *C. paliurus* have dramatically declined and need urgent protection and restoration (Fang et al. 2006). Genome level diversity is still not monitored, indicators that can help identify threats to genetic variation are missing, and there is no strategy for how genetic aspects can be included in biodiversity targets. An improved understanding of *Cyclocarya* genetics would contribute to the formulation of conservation strategies. Here we are reported first complete chloroplast genomes of *Cyclocarya* sequenced to date. The genome was assembled from whole-genome Illumina Hiseq 2500 sequencing data. The annotated chloroplast genomic sequence was deposited into GenBank (Accession Number:KX868928).

Total genomic DNA was extracted from a leaf of *C. paliurus* collected from Guizhou province, China. In this study, voucher specimens of *C. paliurus* were deposited at the herbarium of Northwest University, Xi'an, China and DNA samples were properly stored at Evolutionary Botany Lab, Northwest University, Xi'an, China. After

✉ Peng Zhao
pengzhao@nwu.edu.cn

Yiheng Hu
yihenghu@yeah.net

Jing Yan
734037784@qq.com

Xiaojia Feng
332146704@qq.com

Meng Dang
a1193371050@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, 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

trimming, high-quality PE reads were aligned to the reference (*Juglans regia* chloroplast genome) using bowtie2 v2.2.6 (Langmead and Salzberg 2012), and assembled into a genome using SPAdes v3.60 (Bankevich et al. 2012) assembler. Annotation was performed with Dual Organelar Genome Annotator (DOGMA) software (Wyman et al. 2004). We corrected the annotation with Geneious (Kearse et al. 2012). A map of the genome was generated using OGDRAW (Lohse et al. 2013).

The *C. paliurus* chloroplast genome is 160,562 bp in length. Its organization and gene content were typical of most angiosperms. Overall GC content was 36.1%; GC content in the LSC, SSC and IR regions was 33.7, 29.7 and 42.5 %, respectively. The genome consisted of a large single copy (LSC) region of 900,07 bp, a small single copy (SSC) region of 184,77 bp, and a pair of inverted repeat (IRa and IRb) of 26,039 bp (Fig. 1). It contained 137 unique genes

of which 89 were protein-coding genes, including two copies each of *ndhB*, *rpl2*, *rpl23*, *rps7*, *ycf1*, *ycf15* and *ycf2*, and three copies of *rps12*, two copies of four rRNA genes (4.5 S rRNA, 5 S rRNA, 16 S rRNA, and 23 S rRNA), and 40 tRNA, of which there were two copies of *trnA*-UGC, *trnI*-CAU, *trnI*-GAU, *trnL*-CAA, *trnM*-CAU, *trnN*-GUU, *trnR*-ACG, *trnT*-GGU, and *trnV*-GAC (Fig. 1).

The chloroplast phylogeny of *C. paliurus* plus 11 other Fagales species and *Pyrus pyrifolia* as outgroup was performed based on complete chloroplast genome sequences obtained from the National Center for Biotechnology Information: *Betula nana*: KX703002; *Carya sinensis*: KX671976; *Castanea mollissima*: NC014674; *Castanopsis echinocarpa*: NC_023801; *Corylus fargesii*: NC_031854; *Juglans regia*: KT963008; *Lithocarpus balansae*: KP299291, *Ostrya rehderiana*: NC_028349; *Platycarya strobilacea*: KX868670; *Quercus rubra*: JX970937;

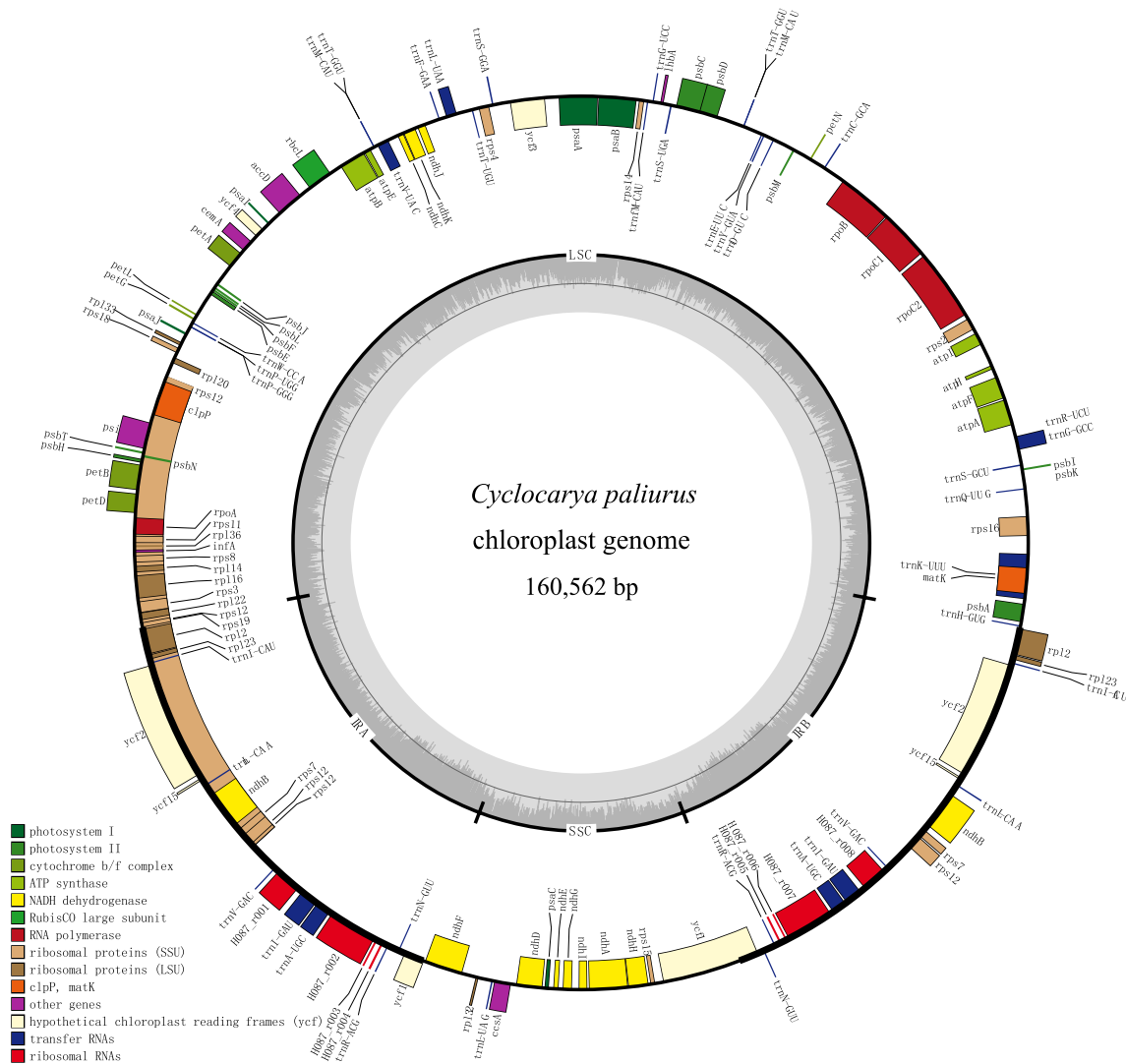
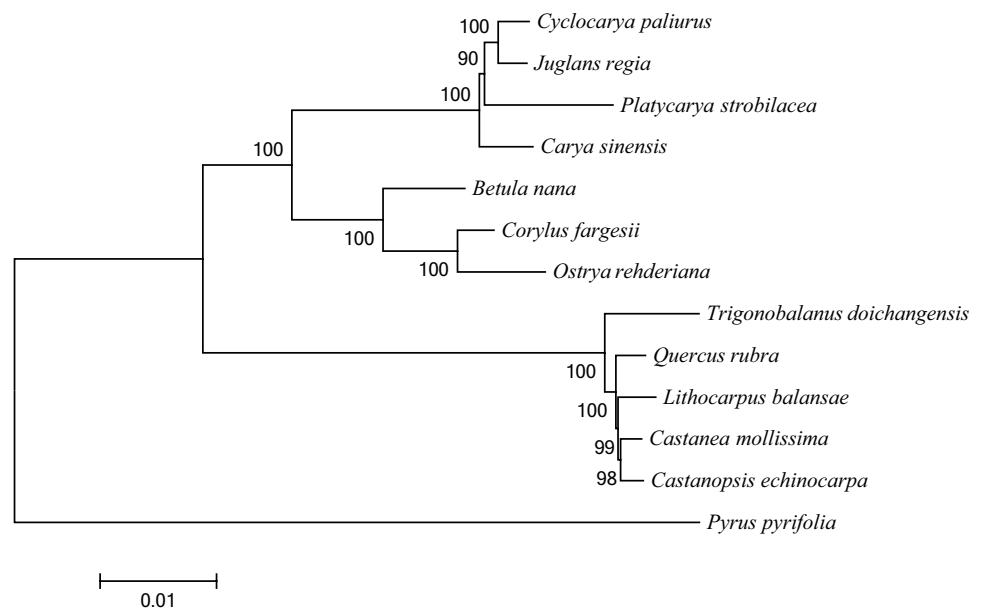


Fig. 1 Physical map of the complete chloroplast genome of *Cyclocarya paliurus*

Fig. 2 Neighbor-joining tree representing chloroplast phylogeny among 13 chloroplast genomes. The 13 species can be divided into three independent clades: Juglandaceae, Betulaceae and Fagaceae. *Pyrus pyrifolia* was used as an out-group. Bootstrap support values are given at the nodes



Trigonobalanus doichangensis: KF990556; *Pyrus pyrifolia*: AP012207. Sequences were aligned using MAFFT v7.017 (Kato and Standley 2013), and the alignment was then manually adjusted. The phylogenetic relationships were inferred using Neighbor-Joining (NJ), which was performed using MEGA v6.0 (Tamura et al. 2013) with the Kimura 2-parameter model. The local bootstrap probability of each branch was calculated by 1000 replications (Fig. 2). The resulting tree showed that *C. paliurus* was most closely related to *Juglans regia*, as expected, with 100% bootstrap support.

Acknowledgements This work was supported by the National Natural Science Foundation of China (41471038, 31200500, and J1210063), the Program for Excellent Young Academic Backbones funding by Northwest University (YZZ15062), 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, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD (2012)

- SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–477
- Crane PR, DuVal A (2013) *Cyclocarya paliurus* (Juglandaceae). *Curtis's Bot Mag* 30:222–232.
- Fang S, Wang J, Wei Z, Zhu Z (2006) Methods to break seed dormancy in *Cyclocarya paliurus* (Batal) Iljinskaja. *Sci Hortic* 110(3):305–309
- 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
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359
- 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
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729
- Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20(17):3252–3255