

# Characterization of the complete chloroplast genome of *Carpinus putoensis*

Shuo Feng<sup>1</sup> · Xiuyue Xie<sup>1</sup> · Mingcheng Wang<sup>1</sup> · Yongzhi Yang<sup>1</sup>

Received: 20 August 2016 / Accepted: 23 August 2016 / Published online: 26 August 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** The whole chloroplast (cp) genome sequence of *Carpinus putoensis* has been characterized from Illumina pair-end sequencing. The complete cp genome was 159,673 bp in length, containing a large single copy region (LSC) of 89,020 bp and a small single copy region (SSC) of 18,568 bp, which were separated by a pair of 26,044 and 26,043 bp inverted repeat regions (IRs). The genome contained 128 genes, including 83 protein-coding genes (75 PCG species), 36 tRNA genes (29 tRNA species) and 8 ribosomal RNA genes (4 rRNA species). The most of gene species occur as a single copy, while 16 gene species occur in double copies. The overall AT content of *C. putoensis* chloroplast genome is 63.6%, while the corresponding values of the LSC, SSC and IR regions are 65.8, 69.8 and 57.5%, respectively. Further, phylogenetic analysis suggested that *C. putoensis* is closely related to *Ostrya rehderiana*.

**Keywords** *Carpinus putoensis* · Chloroplast genome · Phylogenetic analysis

*Carpinus putoensis*, a critically endangered species, is endemic to Putuo Island of the Zhoushan Archipelago, Zhejiang Province, China. The wild *C. putoensis* tree, as

a single mature individual species preserved on the top of Foding Mountain, is under first-grade state protection. This single tree was discovered in the early 1930s and since then no more wild individuals have been found (Shaw et al. 2014). This species is monoecious, in theory it is able to reproduce sexually in the wild, but it has a very low rate of seed production due to strong winds during flowering. Natural regeneration therefore is extremely poor and almost no seedlings are found under the tree. *C. putoensis* is currently listed as a critically endangered species in the IUCN red list (IUCN 2016; Shaw et al. 2014). Therefore, a good knowledge of its genetics would contribute to the formulation of conservation strategy. In this study, we assembled and characterized the complete chloroplast genome sequence of *C. putoensis* based on the Illumina pair-end sequencing data.

Fresh leaves of an individual were collected from Hangzhou Botanical Garden (Zhejiang, China; 30°15'N, 120°16'E), and were used for total genomic DNA extraction with the modified CTAB method (Doyle 1987). The whole-genome sequencing was conducted with 150 bp pair-end reads on the Illumina HiSeq Platform (Illumina, San Diego, CA). In total, about 900 million high quality base pairs (Gb) of sequence data were obtained and used for the cp genome assembly using Velvet software (Zerbino and Birney 2008). The resulting contigs were linked based on overlapping regions after being aligned to *Juglans regia* (NC\_028617) (Peng et al. 2015) and visualized in Geneious version 8.0.5 (Kearse et al. 2012). Annotation was performed with a newly developed command-line application called Plann (Plastome Annotator), which is suitable for annotation of plastomes with a well-annotated sequenced relative (Huang and Cronk 2015). Then, we corrected the annotation with Geneious (Kearse et al. 2012). And next, we generate a physical map of the genome using OGDRAW (<http://ogdraw.mpimp-golm.mpg.de/>) (Lohse et al. 2013).

**Electronic supplementary material** The online version of this article (doi:10.1007/s12686-016-0604-1) contains supplementary material, which is available to authorized users.

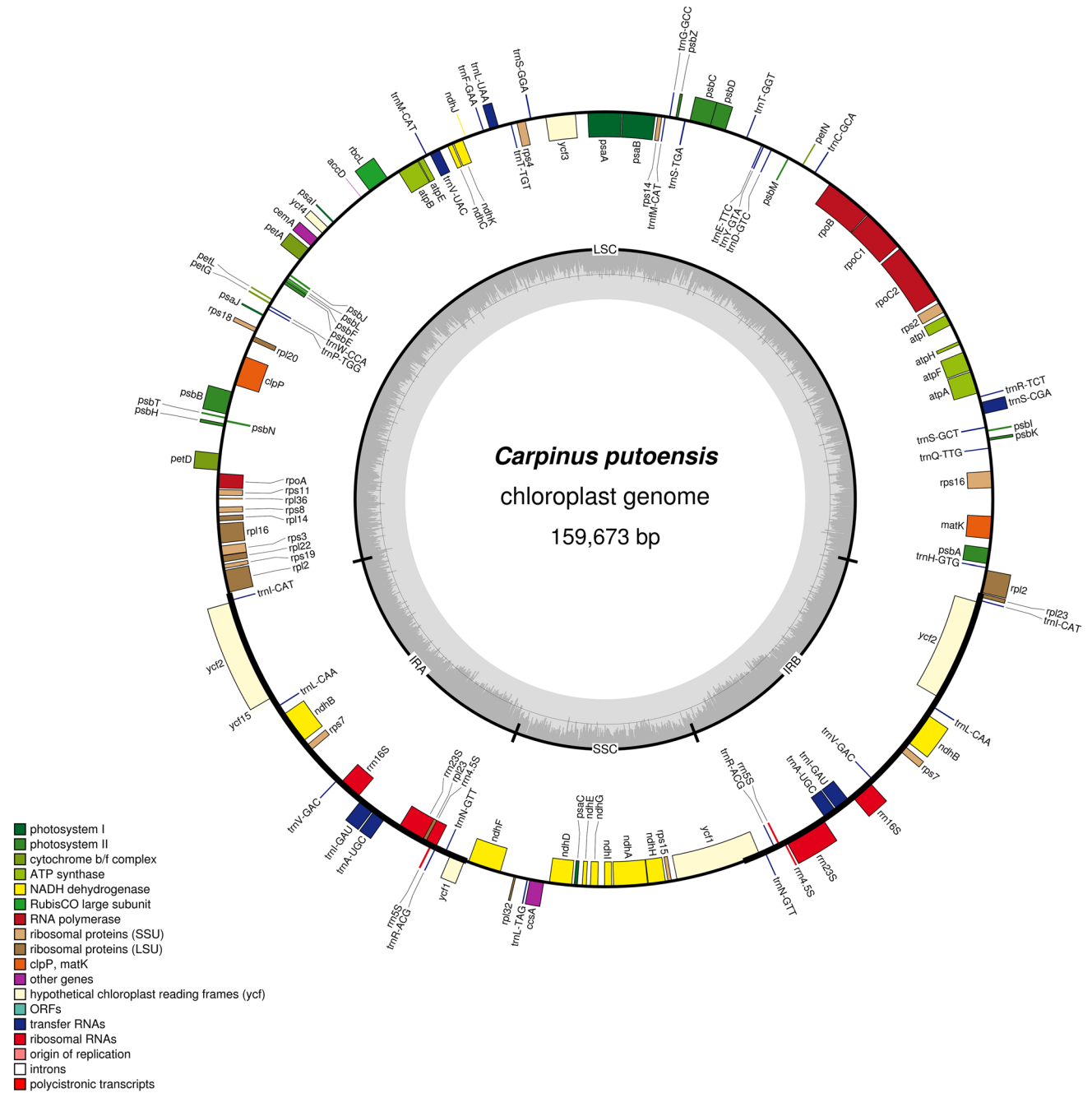
✉ Yongzhi Yang  
yangyongzhi2008@gmail.com

<sup>1</sup> State Key Laboratory of Grassland Agro-Ecosystem, College of Life Science, Lanzhou University, Lanzhou, China

A maximum likelihood (ML) tree with 100 bootstrap replicates was inferred using RaxML version 8 (Stamatakis 2014) from alignments created by the MAFFT (Kato and Standley 2013) using plastid genomes of 10 species (Supplementary Figure S1). The complete chloroplast genome sequence together with gene annotations were submitted to GenBank under the accession numbers of KX695124 for *C. putoensis*.

The complete cp genome of *C. putoensis* is a double stranded, circular DNA 159,673 bp in length, which

contains two inverted repeat (IR) regions of 26,044 and 26,043 bp each separated by a large single-copy (LSC) and a small single-copy (SSC) region of 89,020 and 18,568 bp, respectively (Fig. 1). The genome contained 128 genes, including 83 protein-coding genes, 36 tRNA genes and 8 ribosomal RNA genes. The most of gene species occur as a single copy, while 16 gene species occur in double copies, including all rRNA species (4.5S, 5S, 16S and 23S rRNA), 7 tRNA species (trnA-UGC, trnI-CAT, trnI-GAU, trnL-CAA, trnN-GTT, trnR-ACG and trnV-GAC) and 5 PCG species



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

(*ndhB*, *rpl2*, *rps12*, *rps7* and *ycf2*). The overall AT content of *C. putoensis* cpDNA is 63.6%, while the corresponding values of the LSC, SSC and IR regions are 65.8, 69.8 and 57.5% respectively.

The phylogenetic analysis which *Rosa odorata* var. *gigantea* (Yang et al. 2014) was used as the outgroup showed that the cp genome of *C. putoensis* is closely related to *Ostrya rehderiana* (Supplementary Figure S1). This complete chloroplast genome can be readily used for population genomic studies of *C. putoensis*, and such information would be fundamental to formulate potential new conservation and management strategies for this endangered species.

**Acknowledgments** This work was financially supported by the National High Technology Research and Development Program of China (863 Program, No. 2013AA100605), and International Collaboration 111 Projects of China.

## References

- Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Huang DI, Cronk QCB (2015) Plann: a command-line application for annotating plastome sequences. *Appl. Plant Sci* 3:1500026
- International Union for Conservation of Nature (2016) The IUCN red list of threatened species: *Carpinus putoensis*. doi:10.2305/IUCN.UK.2014-3.RLTS.T32303A2813038.en
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW— a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:W575–W581
- Peng S, Yang G, Liu C, Yu Z, Zhai M (2015) The complete chloroplast genome of the *Juglansregia* (*Juglandales: Juglandaceae*). *Mitochondrial DNA* 1–2. doi:10.3109/19401736.2015.1127367
- Shaw K, Stritch L, Rivers M, Roy S, Wilson B, Govaerts R (2014) The red list of Betulaceae. BGCI, Richmond
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Yang JB, Li DZ, Li HT (2014) Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour* 14:1024–1031
- Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829