

# The complete chloroplast genome sequence of *Corylus chinensis* Franch

Guanglong Hu<sup>1</sup> · Lili Cheng<sup>1</sup> · Yanping Lan<sup>1</sup> · Qingchang Cao<sup>1</sup> · Wugang Huang<sup>1</sup>

Received: 16 September 2016 / Accepted: 6 October 2016 / Published online: 8 October 2016  
© Springer Science+Business Media Dordrecht 2016

**Abstract** This study aimed to map and determine the complete chloroplast genome sequence of a rare Betulaceae species, *Corylus chinensis* Franch., on the basis of the Illumina sequencing data. The complete chloroplast genome is 159,915 bp, and comprises a pair of inverted repeat regions of 26,989 bp each, a large single-copy region of 88,115 bp, and a small single-copy region of 17,822 bp. It harbors 132 genes, including 95 protein-coding genes, 8 rRNA genes, and 29 transfer RNA genes. The overall A+T content of the whole genome is 63.51%. A phylogenetic analysis based on chloroplast genomes has indicated that *C. chinensis* Franch. is closely related to *Ostrya rehderiana*.

**Keywords** Chloroplast genome · *Corylus chinensis* · Illumina sequencing · Phylogenetic analysis

*Corylus chinensis*, the only macrophanerophyte species in the genus *Corylus*, is natively distributed in the southern region of China. As a fruit–timber tree, *C. chinensis* has been seriously threatened by human activities in the last few decades. Recently, several chloroplast (cp) DNA markers have been used for phylogenetic analysis of hazelnut species (Erdogan et al. 2000; Palmé et al. 2002; Boccacci et al. 2009; Bassil et al. 2013; Martins et al. 2013). However, little is known about the chloroplast

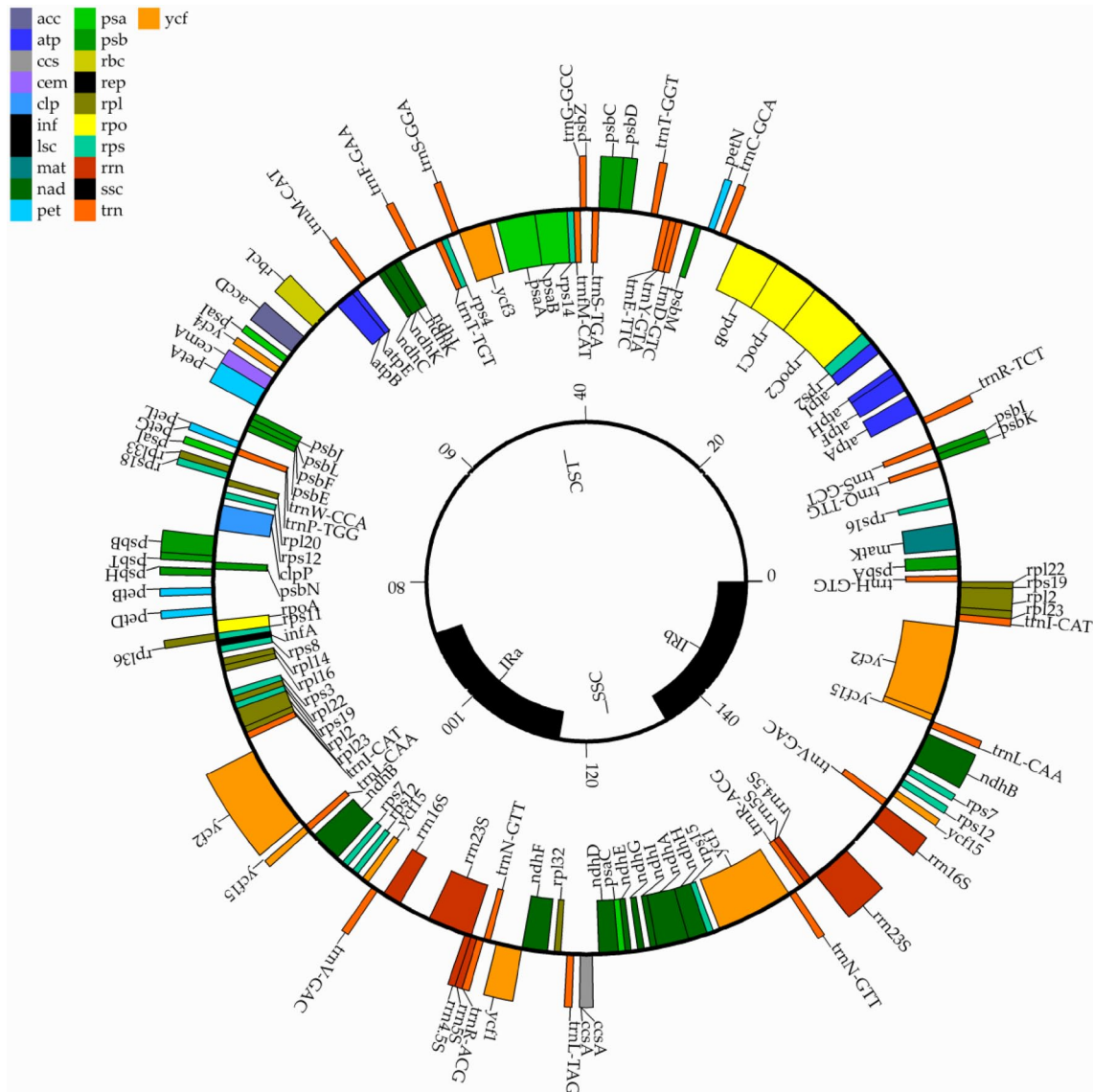
genome of the *Corylus* genus. This study was the first to report the complete chloroplast genome of *C. chinensis* on the basis of the Illumina pair-end sequencing dataset. The annotated chloroplast genome of *C. chinensis* has been deposited into GenBank with the Accession No. KX426347.

Fresh leaves were collected from a single *C. chinensis* plant planted at the germplasm nursery of the Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. DNA extractions were performed using the Plant Genomic DNA Kit (Tiangen Biotech Co., China) according to the manufacturer's instructions, and the high-throughput sequencing was carried out on the Hiseq2500 System (Illumina, USA) by Ori-Gen Technology Co., Beijing, China. In all, 14.55 M of 126-bp raw reads were trimmed with SolexaQA (Cox et al. 2010). About 1.487G high-quality base pairs of sequencing data were obtained and used for the chloroplast genome assembly using SOAPdenovo software (Luo et al. 2012). Then, reference-guided assembly was performed to reconstruct the chloroplast genomes with the program BLAST (Altschul et al. 1990) using closely related species as the references. A 159,915-bp chloroplast genome of *C. chinensis* was obtained after filling the gap with GapCloser (<http://soap.genomics.org.cn/index.html>). The annotation was performed using CPGAVAS (Liu et al. 2012). Finally, a physical map of the chloroplast genome was generated.

The circular genome contains a pair of inverted repeat (IRa and IRb) regions of 26,989 bp, a large single-copy region, and a small single-copy region with the lengths of 88,115 and 17,822 bp, respectively (Fig. 1). The chloroplast genome of *C. chinensis* comprises 132 genes, including 95 protein-coding genes, 8 ribosomal RNA genes, and 29 transfer RNA genes. Among the annotated genes, 11

✉ Wugang Huang  
huang\_wugang@hotmail.com

<sup>1</sup> Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China



**Fig. 1** Gene map of the *Corylus chinensis* chloroplast genome

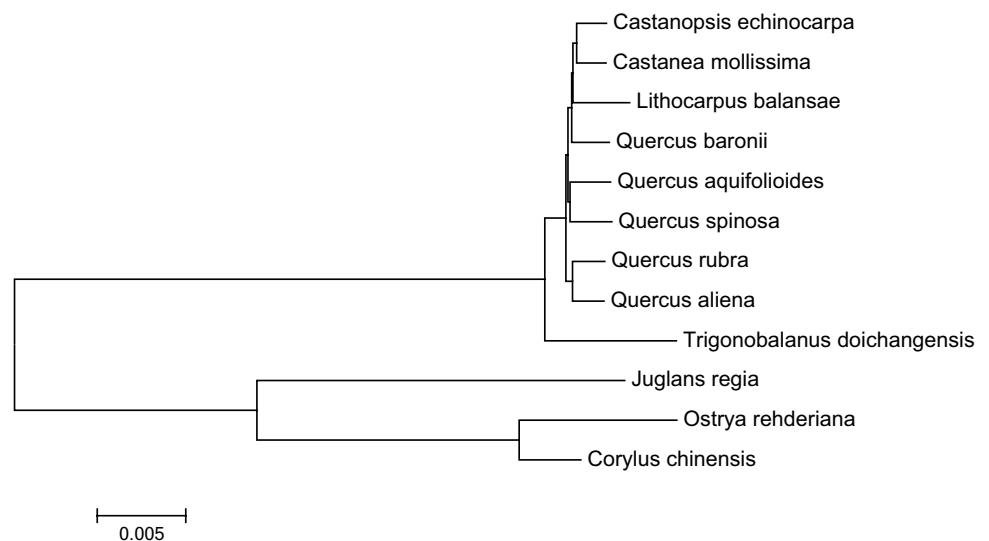
protein-coding genes contain introns, including 9 genes with a single intron each (*atpF*, *rpoC1*, *rpl2*, *ycf15*, *ndhB*, *ndhA*, *ndhB*, *ycf15*, and *rpl2*) and 2 genes with 2 introns each (*clpP* and *ycf3*). The overall AT content of *C. chinensis* cp genome is 63.51 %.

On the basis of the cpDNA sequences of *C. chinensis* and the other 11 species of the families *Fagaceae*, *Juglandaceae*, and *Betulaceae* from the Fagales, a neighbor-joining phylogenetic tree was conducted using MEGA6

(Tamura et al. 2013). The cpDNA sequences of 12 plants were analyzed to identify the phylogenetic position of *C. chinensis* using MEGA6. The phylogenetic analysis showed that the cp genome of *C. chinensis* was closely related to that of *Ostrya rehderiana* of the family *Betulaceae* (Fig. 2).

The aforementioned complete chloroplast genome can be subsequently used for population, phylogenetic, and cp genetic engineering studies of *C. chinensis*, and such

**Fig. 2** Neighbor-joining phylogenetic tree inferred using MEGA6 from 12 complete cp genomes



information would be fundamental to formulate potential new conservation and management strategies for this important *Betulaceae* species.

## References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
- Bassil N, Boccacci P, Botta R, Postman J, Mehlenbacher S (2013) Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars. *Genet Resour Crop Evol* 60:543
- Boccacci P, Botta R (2009) Investigating the origin of hazelnut (*Corylus avellana* L.) cultivars using chloroplast microsatellites. *Genet Resour Crop Evol* 56:851–859
- Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform* 11:485
- Erdogan V, Mehlenbacher S (2000) Phylogenetic relationships of *Corylus* species (*Betulaceae*) based on nuclear ribosomal DNA ITS region and chloroplast matK gene sequences. *Syst Bot* 25(4):727–737
- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X (2012) CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genom* 13(1):715
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Zhu X, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1(1):1
- Martins S, Simões F, Mendonça D, Matos J, Silva AP, Carnide V (2013) Chloroplast SSR genetic diversity indicates a refuge for *Corylus avellana* in northern Portugal. *Genetic Resour Crop Evol* 60(4):1289–1295
- Palmé AE, Vendramin GG (2002) Chloroplast DNA variation, postglacial recolonization and hybridization in hazel, *Corylus avellana*. *Mol Ecol* 11:1769–1779
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. doi:10.1093/molbev/mst197