Conservation Genet Resour (2017) 9:225–227
DOI 10.1007/s12686-016-0656-2

TECHNICAL NOTE

The complete chloroplast genome sequence of the endangered Chinese endemic tree *Corylus fargesii*

Guanglong Hu[1] · Lili Cheng[1] · Yanping Lan[1] · Qingchang Cao[1] · Xueqiang Wang[2] · Wugang Huang[1]

Received: 19 October 2016 / Accepted: 16 November 2016 / Published online: 21 November 2016

Abstract The complete chloroplast genome sequence of a Betulaceae species, *Corylus fargesii*, was mapped and determined based on Illumina sequencing data. The complete chloroplast genome was shown to be 159,856 bp, and comprises a pair of inverted repeat regions of 26,602 bp each, a large single-copy region of 88,313 bp, and a small single-copy region of 18,339 bp. It harbors 131 genes, including 94 protein-coding genes, 8 ribosomal RNA genes, and 29 transfer RNA genes. The A+T content of the entire genome is 63.49%. Phylogenetic analysis based on chloroplast genomes indicates that *C. fargesii* is closely related to *Ostrya rehderiana*.

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

Corylus fargesii, a Chinese endemic hazelnut species, grows in the mountain valleys from the elevation of 800–3000 m of the southwest China (Kuang et al. 1979). As a small arbor species in the genus *Corylus, C. fargesii* has been seriously threatened by human activities in the last few decades. Although several chloroplast (cp) DNA markers have previously been used for phylogenetic analysis (Erdogan and Mehlenbacher 2000; Palmé and Vendramin 2002; Boccacci and Botta 2009; Bassil et al. 2013; Martins et al. 2013), except the *Corylus Chinensis* (GenBank accession number KX814336, Hu et al. 2016), little is known about the chloroplast genome in this genus. In the present study, we report the first complete chloroplast genome of *C. fargesii* based on Illumina pair-end sequencing data. The annotated chloroplast genome of *C. fargesii* has been deposited into GenBank under accession number KX822767.

Fresh leaves were collected from a single *C. fargesii* plant that grew at the resources nursery of the Institute of Forestry and Pomology (Beijing Academy of Agriculture and Forestry Sciences, Beijing, China). DNA extraction was performed using the Plant Genomic DNA Kit (TIANGEN Biotech Co., Beijing, China) according to the manufacturer's instructions, and high-throughput sequencing was carried out on the HiSeq2500 System (Illumina, San Diego, CA) by OriGene Technologies (Beijing, China). A total of 80.43 M of 126-bp raw reads were trimmed using SolexaQA (Cox et al. 2010), and approximately 9.577 G high-quality base pairs of sequence data were obtained and used for chloroplast genome assembly with SOAPdenovo software (Luo et al. 2012). Reference-guided assembly was then performed to reconstruct the chloroplast genome with the BLAST program (Altschul et al. 1990) using closely related species as references. After filling the gaps with GapCloser (http://soap.genomics.org.cn/index.html), a 159,856 bp chloroplast genome of *C. fargesii* was obtained. Annotation was performed using the CpGAVAS web server (Liu et al. 2012) to generate a physical map of the chloroplast genome.

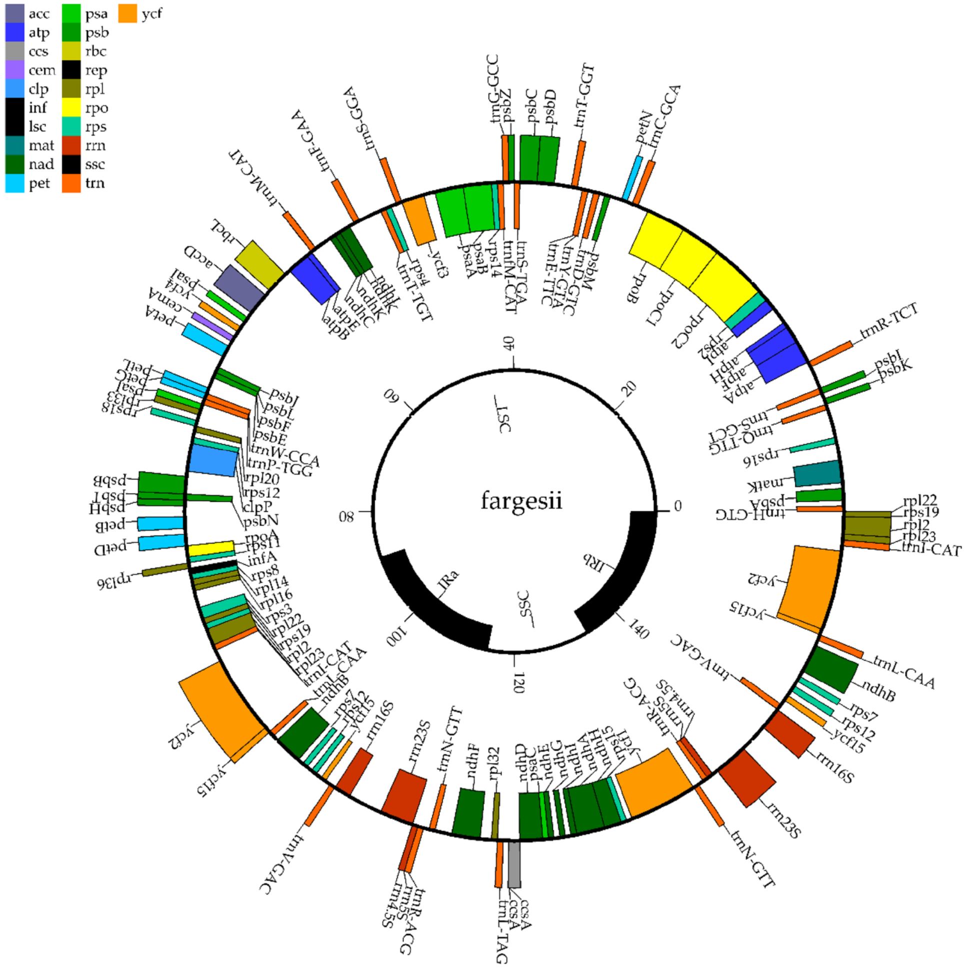
The circular chloroplast genome of *C. fargesii* contains a pair of inverted repeat (IRa and IRb) regions, each of 26,602 bp, a large single-copy (LSC) region and a small single-copy (SSC) region of 88,313 and 18,339 bp, respectively (Fig. 1). It comprises 131 genes, including 94

✉ Wugang Huang
huang_wugang@hotmail.com

[1] Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China

[2] College of Agronomy and Biotechnology, China Agricultural University, Beijing, China

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



protein-coding genes, 8 ribosomal RNA genes, and 29 transfer RNA genes. Among the annotated genes, 11 protein-coding genes contain introns, including nine with a single intron each (*atpF, rpoC1, rpl2, ycf15, ndhB, ndhA, ndhB, ycf15*, and *rpl2*), and two with two introns each (*clpP* and *ycf3*). The overall AT content of the *C. fargesii* cp genome is 63.49%.

To identify the phylogenetic position of *C. fargesii*, maximum likelihood analysis was performed based on the 12 Fagales plant chloroplast genomes and seven other outgroup plants using MEGA6 software (Tamura et al. 2013). The cp genome of *C. fargesii* was shown to be closely related to that of *Ostrya rehderiana* of the family Betulaceae (Fig. 2).

This complete chloroplast genome can be used for subsequent population, phylogenetic, and cp genetic engineering studies of *C. fargesii*. The new information will also be fundamental in formulating potential new conservation and management strategies for this important hazel species.

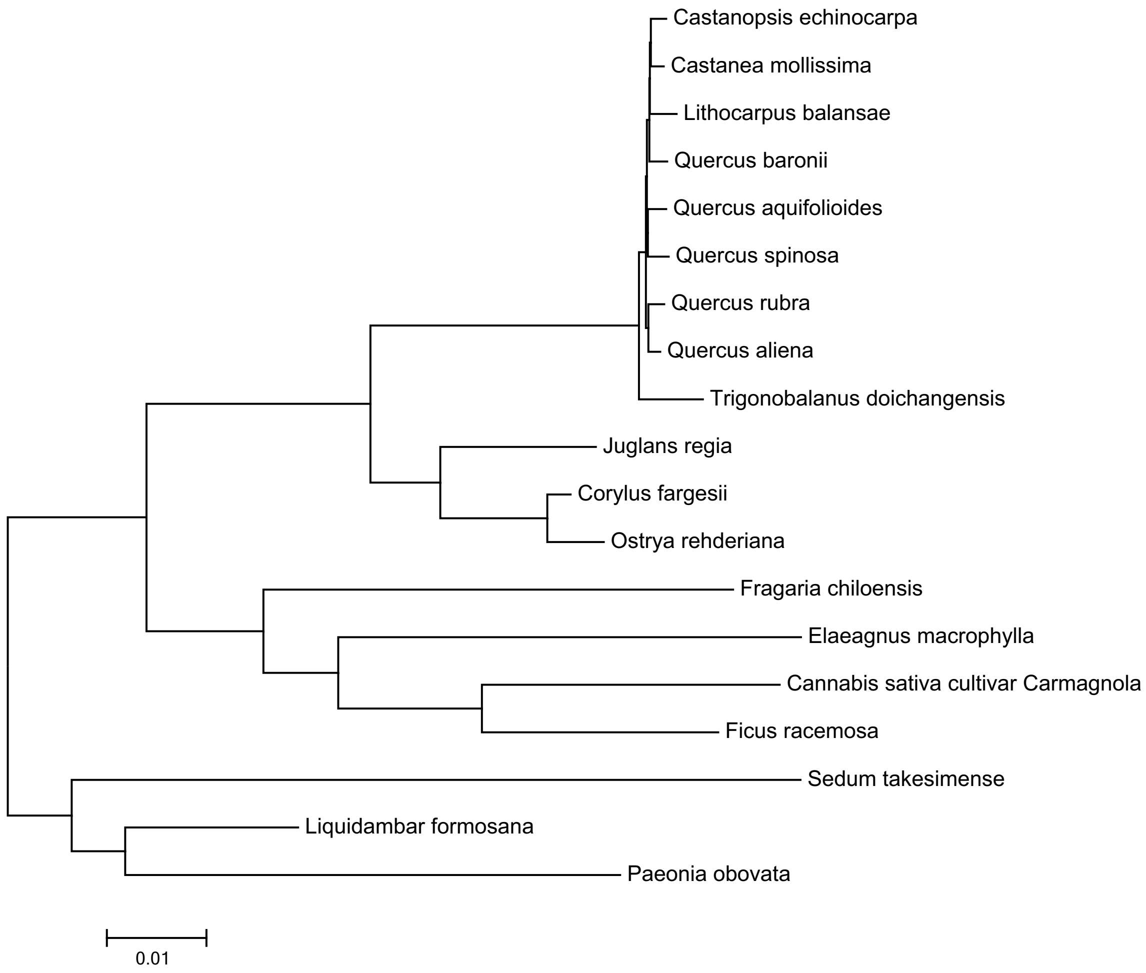


Fig. 2 Phylogenetic tree inferred using MEGA6 software from 19 complete cp genomes

Acknowledgements This work was financially supported by Young Foundation of Beijing Academy of Agriculture and Forestry Sciences (QNJJ201405), Beijing talents training Program (2014000020060G179), Special Fund for the Construction of Scientific and Technological Innovation Capability (KJCX20140110).

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 Illuminasecond-generation sequencing data. BMC Bioinformatics 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

Hu G, Cheng L, Lan Y, Cao Q, Huang W (2016) The complete chloroplast genome sequence of *Corylus chinensis* Franch. Conserv Genet Resour. doi:10.1007/s12686-016-0636-6

Kuang KR, Zheng SX, Li PQ, Lu AM (1979) Flora of China vol. 21, Science Press, Beijing. p 54

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 Genomics 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. Genet 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, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729. doi:10.1093/molbev/mst197