

## The complete chloroplast genome sequence of an economic plant *Coffea canephora*

Dongyang Wu<sup>a,b</sup>, Changwei Bi<sup>a</sup>, Xuelin Wang<sup>a</sup>, Yiqing Xu<sup>a</sup>, Qiaolin Ye<sup>a</sup> and Ning Ye<sup>a</sup> 

<sup>a</sup>College of Information Science and Technology, Nanjing Forestry University, Nanjing, Jiangsu, China; <sup>b</sup>College of Forest Resources and Environment, Nanjing Forestry University, Nanjing, Jiangsu, China

### ABSTRACT

*Coffea canephora* is a paramount economic plant with great economic value. The complete chloroplast (cp) genome is 154,751 bp in length, including a large single copy (LSC) region of 84,850 bp, a small single copy (SSC) region of 18,131 bp and a pair of inverted repeats (IRs) of 25,885 bp. This cp genome contains 131 genes, comprising of 86 protein-coding genes, 37 tRNAs and 8 rRNAs. The majority of these genes were single copy genes, while 18 genes existed as double copies, including 6 protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7* and *ycf2*), 8 tRNA genes (*trnA-UGC*, *trnG-GCC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG* and *trnV-GAC*) and 4 rRNA genes (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*). A neighbour-joining phylogenetic tree was reconstructed to indicate that *Coffea canephora* is evolutionarily close to *Olea europaea* within Asterids. The complete cp genome will provide intragenic information for molecular phylogeny and biological studies of the Rubiales.

### ARTICLE HISTORY

Received 30 June 2017  
Accepted 26 July 2017

### KEYWORDS

*Coffea canephora*;  
chloroplast genome;  
rubiaceae; phylogeny

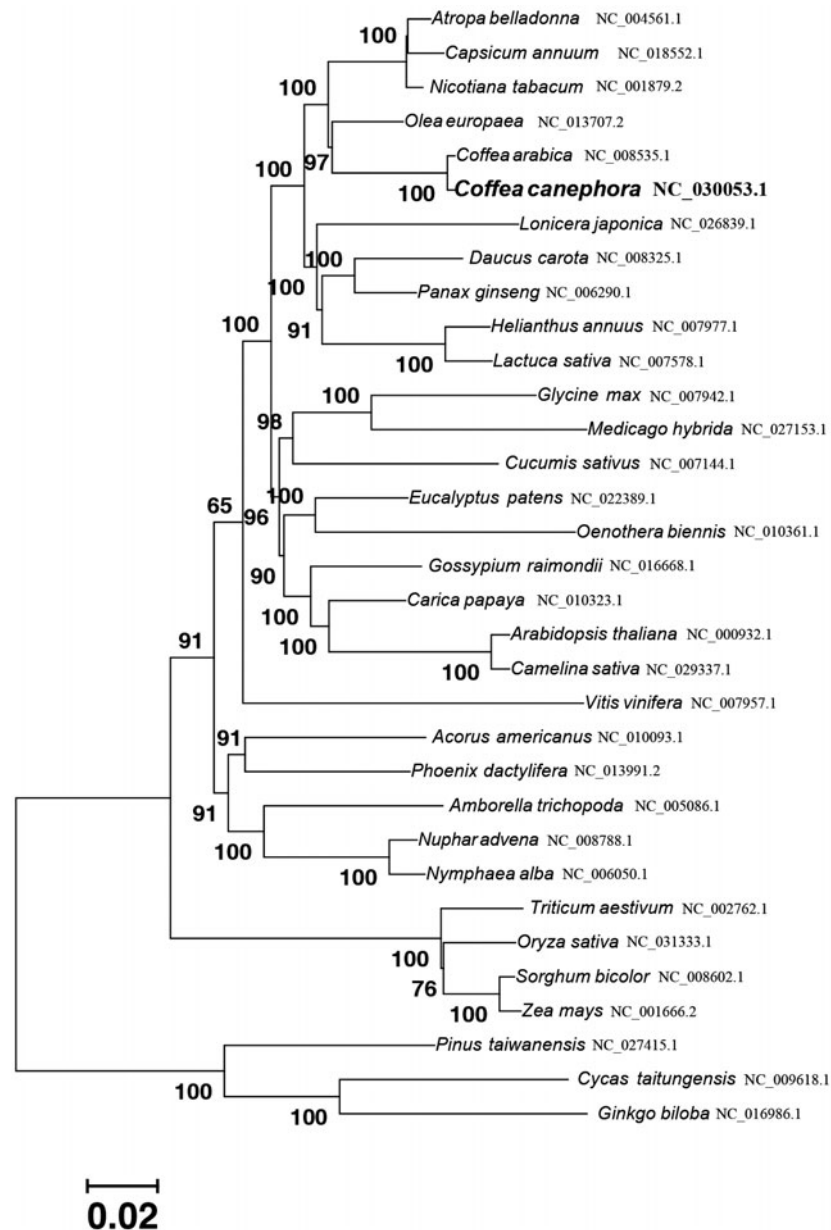
*Coffea canephora* is a member of Rubiaceae, the fourth largest family of angiosperms. The genus *Coffea* comprises more than one hundred species but only about 25 produce fruits with commercial value. Additionally, from about 90 species of *Coffea*, only *Coffea arabica* (Arabica coffee) and *Coffea canephora* (Robusta coffee) are significant in the world commercial trade (Damatta et al. 2003). *Coffea canephora* is more productive and resistant to diseases (Servillo et al. 2016). *Coffea canephora* is a cultivated plant of economic prominence that is the second most heavily traded commodity worldwide (Morais et al. 2012). This crop is second in value only to oil as a source of foreign exchange to several developing countries.

Chloroplasts are essential organelles in plants that originated from Cyanobacteria by endosymbiosis with the precursor of nucleated ancestral cells more than 1.2 billion years ago (Timmis et al. 2004). The majority of angiosperm cp genomes are highly conserved in gene contents and order (Wicke et al. 2011). Since the first complete chloroplast (cp) genome sequence of liverwort (*Marchantia polymorpha*) reported in 1986 (Ohyama et al. 1986), almost 1400 cp genomes have been deposited in NCBI Organelle Genome Resources database. In this study, we described the assembly and annotation details of the *Coffea canephora* cp genome (NCBI Accession Number: NC\_030053.1), which will provide a convenient tool in the study of molecular identification, genetic diversity and phylogenetic classification in Rubiales.

Young expanding leaves from two greenhouse grown *Coffea canephora* at IRD Montpellier (Montpellier, France) were harvested and stored at  $-80^{\circ}\text{C}$  prior to DNA extraction.

A large quantity of genomic DNA was extracted by means of a nuclei isolation step as described by Carrier et al. (2011) and then deposited at Energy and Sustainable Development (ENEA), Casaccia Research Center of Italy. All data were generated using next-generation sequencers (Roche/454 GSFLX and Illumina GA IIx) (Denoeud et al. 2014). The original sequencing reads were a mixture of DNA with nucleus and other organelles. The sequences were first assembled using Newbler 3.0 (454-Roche) with default parameters. In order to filter cp contigs, we researched five similar plants' cp genome sequences (*Coffea arabica*, *Capsicum annum*, *Nicotiana tabacum*, *Olea europaea* and *Ginkgo biloba*) from NCBI Genbank, and then we used BLASTN to isolate cp contigs based on these complete reference cp sequences (Camacho et al. 2009). To visualize the contigs connections, we used Perl scripts and Newbler 3.0 generated file '454AllContigGraph.txt' (Bi et al. 2016). Gaps between contigs were filled up with the method described by Ma et al. (2017). Finally, a 154,751 bp nucleotide genome sequence was finished and annotated with DOGMA (Wyman et al. 2004). The physical map of *Coffea canephora* cp genome was generated with OGDRAW (Lohse et al. 2007).

The complete *Coffea canephora* cp genome is 154,751 bp in length, exhibiting a typical quadripartite structure of a LSC region of 84,850 bp, a SSC region of 18,131 bp and a pair of IR region of 25,885 bp. The overall base composition of the cp genome in asymmetric order is A: 30.92%, C: 19.09%, G: 18.38% and T: 31.61%, and the AT content is 62.53%. The proportions of AT contents in LSC, SSC and IR regions are



**Figure 1.** Phylogenetic relationships based on amino acid sequences of the 77 protein-coding genes of 33 species. These conserved genes were aligned with ClustalW and the phylogenetic tree was constructed using the neighbour-joining method in MEGA6. Bootstrapping values are listed for each node.

64.6%, 68.71% and 56.95%, respectively. Using the online program DOGMA, a total of 131 genes were identified in the cp genome, including 86 protein-coding genes, 37 tRNAs and 8 rRNAs. The majority of these genes were single copy genes, whereas 18 genes existed as double copies, including 6 protein-coding genes (*ndhB*, *rpl2*, *rpl23*, *rps12*, *rps7* and *ycf2*), 8 tRNA genes (*trnA-UGC*, *trnG-GCC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG* and *trnV-GAC*) and 4 rRNA genes (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*). Additionally, a total of 11 protein-coding genes were found to have one (*rps16*, *rpl16*, *rpoC1*, *ndhA*, *petB* and *atpF*) or two (*rps12*, *rpl2*, *ndhB*, *ycf3* and *clpP*) introns. To confirm the phylogenetic position of *Coffea canephora* cp genome, 77 protein-coding genes commonly present were extracted from cp genomes of 39 species to reconstruct the neighbour-joining phylogenetic tree, and three gymnosperm species (*Pinus taiwanensis*, *Cycas*

*taitungensis* and *Ginkgo biloba*) were set as out-groups. Through the phylogenetic tree (Figure 1), the cp genome of *Coffea canephora* is evolutionarily close to that of *Coffea arabica* and *Olea europaea*, then *Nicotiana tabacum*, *Capsicum annuum* and *Atropa belladonna*.

## Acknowledgments

We would like to acknowledge Victor A. Albert, Patrick Wincker and Philippe Lashermes for their assistance in sampling material of *Coffea canephora*.

## Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## Funding

This study was supported by the National Key Research and Development Plan of China (2016YFD0600101), the Fundamental Research Funds for the Central Non-Profit Research Institution of CAF (CAFYBB2014QB015), the National Natural Science Foundation of China (31570662, 31500533, and 61401214), Jiangsu Provincial Department of Housing and Urban-Rural Development (2016ZD44), 2017 Graduate Research and Innovation Program Projects in Jiangsu Province (KYCY17\_0827) and the PAPD (Priority Academic Program Development) program at Nanjing Forestry University.

## ORCID

Ning Ye  <http://orcid.org/0000-0001-7249-8352>

## References

- Bi C, Wang X, Xu Y, Wei S, Shi Y, Dai X, Yin T, Ye N. 2016. The complete mitochondrial genome of *Medicago truncatula*. *Mitochondrial Dna*. 1:122-123.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10:421.
- Carrier G, Santoni S, Rodier-Goud M, Canaguier A, Ad K, Dubreuil-Tranchant C, This P, Boursiquot JM, Le CL. 2011. An efficient and rapid protocol for plant nuclear DNA preparation suitable for next generation sequencing methods. *Am J Bot*. 98:e13.
- Damatta FM, Chaves ARM, Pinheiro HA, Ducatti C, Loureiro ME. 2003. Drought tolerance of two field-grown clones of *Coffea canephora*. *Plant Sci*. 164:111-117.
- Denoeud F, Carreteropaulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*. 345:1181-1184.
- Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet*. 52:267-274.
- Ma Q, Li S, Bi C, Hao Z, Sun C, Ye N. 2017. Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). *Curr Genet*. 63:117-129.
- Morais LE, Cavatte PC, Detmann KC, Sanglard LMVP, Ronchi CP, Damatta FM. 2012. Source strength increases with the increasing precociousness of fruit maturation in field-grown clones of conilon coffee (*Coffea canephora*) trees. *Trees*. 26:1397-1402.
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*. 322:572-574.
- Servillo L, Giovane A, Casale R, Cautela D, D'Onofrio N, Balestrieri ML, Castaldo D. 2016. Homostachydrine (pipercolic acid betaine) as authentication marker of roasted blends of *Coffea arabica* and *Coffea canephora* (Robusta) beans. *Food Chem*. 205:52.
- Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet*. 5:123-135.
- Wicke S, Schneeweiss GM, Depamphilis CW, Kai FM, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol*. 76:273-297.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organelle genomes with DOGMA. *Bioinformatics*. 20:3252-3255.