





The complete chloroplast genome of coffee tree, *Coffea arabica* L. 'Typica' (Rubiaceae)

Juhyeon Min^{a,b} , Yongsung Kim^{a,b} , Hong Xi^{a,b}, Kyeong-In Heo^{a,b}  and Jongsun Park^{a,b} 

^aInfoBoss Co., Ltd., Seoul, Republic of Korea; ^bInfoBoss Research Center, Seoul, Republic of Korea

ABSTRACT

We presented complete chloroplast genome of *Coffea arabica* L. 'Typica' which is 155,187 bp long and has four subregions: 85,159 bp of large single copy (LSC) and 18,135 bp of small single copy (SSC) regions are separated by 25,946 bp of inverted repeat (IR) regions including 131 genes (86 protein-coding genes, eight rRNAs, and 37 tRNAs). The overall GC content of the chloroplast genome is 37.4% and those in the LSC, SSC, and IR regions are 35.4%, 31.3%, and 43.0%, respectively. Interestingly, almost most of sequence variations identified against five coffee chloroplast genomes are insertions and deletions.

ARTICLE HISTORY

Received 8 May 2019
Accepted 17 May 2019

KEYWORDS

Coffea arabica; Typica; chloroplast genome; low-level sequence variations

Coffea arabica L. 'Typica' is one of the major coffee cultivars, which is the first coffee varietal with Bourbon group (Philippe et al. 2009). Since *C. arabica* was taken from Southern Ethiopia to Yemen, it had been cultivated for centuries (Meyer 1965). After spreading into Southeast Asia, a single plant from Indonesia had been cultivated in Amsterdam, which originated in cultivar Typica (Anthony et al. 2001). Nowadays it has been cultivated in 'Coffee Belt' (Anthony et al. 2002; Park, Kim, Xi, Nho, et al. 2019). Though Typica is sensitive to several coffee diseases (e.g. coffee leaf rust and coffee berry disease; Van der Vossen 2009), it provides excellent quality widely. We selected Typica for investigation intra-species variation on coffee chloroplast genome.

Total DNA of coffee tree bought in Coffee Cupper, Korea (TY1; YS. Kim, IB-01022, InfoBoss Cyber Herbarium (IN)) was extracted from fresh leaves by using DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). Sequencing was performed using HiSeq2000 at Macrogen Inc., Korea. Raw reads were filtered by Trimmomatic 0.33 (Bolger et al. 2014). *de novo* assembly and sequence confirmation were done by Velvet 1.2.10 (Zerbino and Birney 2008), SOAPGapCloser 1.12 (Zhao et al. 2011), BWA 0.7.17 (Li 2013), and SAMtools 1.9 (Li et al. 2009). Geneious R11 11.0.5 (Biomatters Ltd., Auckland, New Zealand) was used for chloroplast genome annotation based on *Coffea arabica* CH3 chloroplast genome (MK342634; Park, Xi, et al. 2019).

TY1 chloroplast genome (Genbank accession is MK862266) is 155,187 bp long (GC ratio is 37.4%) and has four subregions: 85,159 bp of large single copy (35.4%) and 18,135 bp

of small single copy (31.3%) regions are separated by 25,946 bp of inverted repeat (IR; 43.0%). It contains 131 genes (86 protein-coding genes, eight rRNAs, and 37 tRNAs); 19 genes (8 protein-coding genes, 4 rRNAs, and 7 tRNAs) are duplicated in IR regions.

Based on alignment with five *C. arabica* chloroplasts, CR (Samson et al. 2007), CA (KY085909), CH3 (Park, Xi, et al. 2019), HP1 (Park, Kim, Xi, Nho, et al. 2019), and IN1 (Park, Kim, Xi, Heo 2019), eight insertions and deletions (INDELS) and three single nucleotide polymorphisms (SNPs), three INDELS, four INDELS, four INDELS, and 90 INDELS were found, respectively. Most of sequence variations on coffee chloroplast genomes are INDELS, which is similar to those of *Marchantia polymorpha* subsp. *ruderalis* (69 SNPs and 660 INDELS; Kwon et al. 2019), *Illicium anisatum* (21 SNPs and 114 INDELS; Park, Kim, Xi 2019), *Camellia japonica* (78 SNPs and 643 INDELS; Park, Kim, Xi, Oh, et al. 2019), *Pseudostellaria palibiniana* (84 SNPs and 125 INDELS; Kim et al. 2019), and *Pyrus ussuriensis* (121 SNPs and 781 INDELS; Cho et al. under review).

Seven *Coffea* and four chloroplast genomes of Rubiaceae as outgroup species were used for constructing bootstrapped neighbor joining and maximum likelihood trees using MEGA X (Kumar et al. 2018) based on alignment by MAFFT 7.388 (Kato and Standley 2013). Phylogenetic trees show that TY1 is clustered tightly with other *Coffea arabica* chloroplasts (Figure 1). This chloroplast genome presents low-level of genetic diversity of *C. arabica* chloroplast genomes.

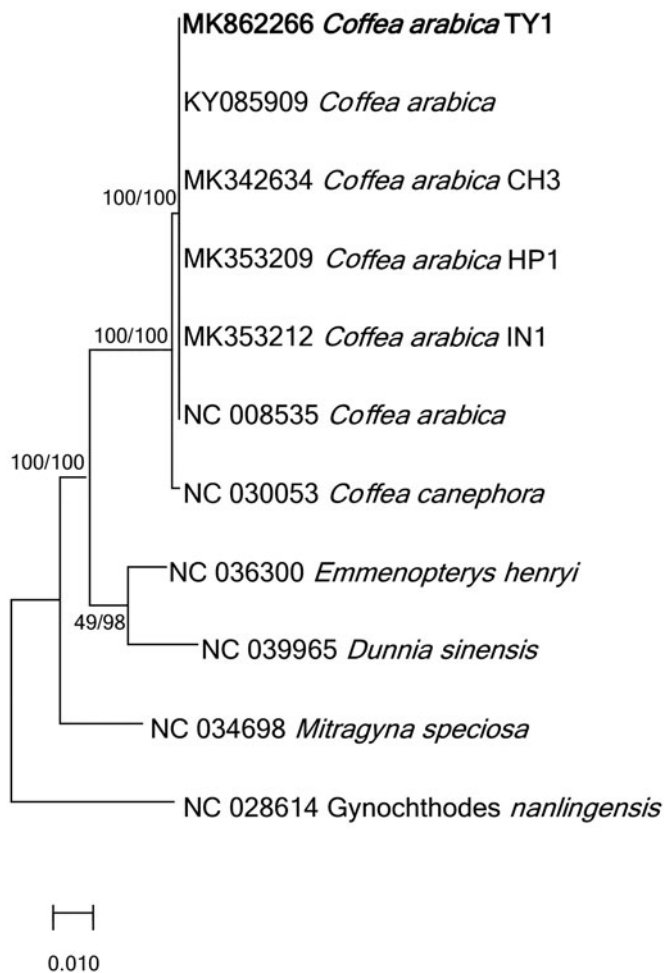


Figure 1. Neighbor joining (bootstrap repeat is 10,000) and maximum-likelihood (bootstrap repeat is 1,000) phylogenetic trees of six *Coffea* and four Rubiaceae complete chloroplast genomes: six *Coffea arabica* (MK862266 in this study, MK353212, NC_008535, KY085909, MK342634, and MK353209), *Coffea canephora* (NC_030053), *Mitragyna speciosa* (NC_034698), *Dunnia sinensis* (NC_039965), *Emmenopterys henryi* (NC_036300), and *Gynochthodes nanlingensis* (NC_028614). Phylogenetic tree was drawn based on maximum likelihood tree. The numbers above branches indicate bootstrap support values of maximum likelihood and neighbor joining phylogenetic trees, respectively.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by InfoBoss Research Grant [IBG-0011].

ORCID

Juhyeon Min  <http://orcid.org/0000-0002-9466-5681>
 Yongsung Kim  <http://orcid.org/0000-0002-5349-9226>
 Kyeong-In Heo  <http://orcid.org/0000-0002-6951-8102>
 Jongsun Park  <http://orcid.org/0000-0003-0786-4701>

References

- Anthony F, Bertrand B, Quiros O, Wilches A, Lashermes P, Berthaud J, Charrier A. 2001. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica*. 118:53–65.
- Anthony F, Combes M, Astorga C, Bertrand B, Graziosi G, Lashermes PJT, Genetics A. 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theor Appl Genet*. 104:894–900.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–2120.
- Cho M-S, Kim Y, Kim S-C, Park J. under review. The complete chloroplast genome of Korean *Pyrus ussuriensis* Maxim. (Rosaceae): providing genetic background of two types of *P. ussuriensis*. doi:10.1080/23802359.2019.1598802
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
- Kim Y, Heo K-I, Park J. 2019. The second complete chloroplast genome sequence of *Pseudostellaria palibiniana* (Takeda) Ohwi (Caryophyllaceae): intraspecies variations based on geographical distribution. *Mitochondrial DNA B*. 4:1310–1311.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35:1547–1549.
- Kwon W, Kim Y, Park J. 2019. The complete chloroplast genome of Korean *Marchantia polymorpha* subsp. *ruderalis* Bischl. & Boisselier: low genetic diversity between Korea and Japan. *Mitochondrial DNA B*. 4:959–960.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Preprint arXiv*. 13033997.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25:2078–2079.
- Meyer FG. 1965. Notes on wild *Coffea arabica* from Southwestern Ethiopia, with some historical considerations. *Econ Bot*. 19:136–151.
- Park J, Kim Y, Xi H. 2019. The complete chloroplast genome of aniseed tree, *Illicium anisatum* L. (Schisandraceae). *Mitochondrial DNA B*. 4: 1023–1024.
- Park J, Kim Y, Xi H, Heo K-I. 2019. The complete chloroplast genome of ornamental coffee tree, *Coffea arabica* L. (Rubiaceae). *Mitochondrial DNA B*. 4:1059–1060.
- Park J, Kim Y, Xi H, Nho M, Woo J, Seo Y. 2019. The complete chloroplast genome of high production individual tree of *Coffea arabica* L. (Rubiaceae). *Mitochondrial DNA B*. 4:1541–1542.
- Park J, Kim Y, Xi H, Oh YJ, Hahm KM, Ko J. 2019. The complete chloroplast genome of common camellia tree, *Camellia japonica* L. (Theaceae), adapted to cold environment in Korea. *Mitochondrial DNA B*. 4:1038–1040.
- Park J, Xi H, Kim Y, Heo K-I, Nho M, Woo J, Seo Y, Yang JH. 2019. The complete chloroplast genome of cold hardiness individual of *Coffea arabica* L. (Rubiaceae). *Mitochondrial DNA B*. 4:1083–1084.
- Philippe L, Benoît B, Hervé E. 2009. Breeding coffee (*Coffea arabica*) for sustainable production. In Jain SM, Priyadarshan PM, editors. Breeding plantation tree crops: tropical species. New York: Springer; p. 525–543.
- Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H. 2007. The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. *Plant Biotechnol J*. 5:339–353.
- Van der Vossen H. 2009. The cup quality of disease-resistant cultivars of Arabica coffee (*Coffea arabica*). *Exp Agric*. 45:323–332.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res*. 18:821–829.
- Zhao Q-Y, Wang Y, Kong Y-M, Luo D, Li X, Hao P. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*. 12:52.