

Comparative analysis of chloroplast genome sequences of four camellia species

Hao Bingqing

Guangxi Forestry Research Institute

Xia Yingying

Guangxi Forestry Research Institute

Zhang Zhaoyuan

Guangxi Forestry Research Institute

Wang Dongxue

Guangxi Forestry Research Institute

Ye hang (✉ gfri_yehang@163.com)

Guangxi Forestry Research Institute

Ma Jinlin

Guangxi Forestry Research Institute

Research Article

Keywords: Camellia osmantha cv'yidan', Chloroplast genome, Genetic structure, Phylogenetic analysis.

Posted Date: May 23rd, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2836580/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Researching the photosynthetic characteristics based on the whole chloroplast genome sequence of *Camellia osmantha* cv 'yidan' is important for improving production. We sequenced and analyzed the chloroplast (cp) genomes of *C. osmantha* cv 'yidan'. The total cp genome length was 156,981 bp. The cp genomes included 134 genes encoding 81 proteins, 39 transfer RNAs, 8 ribosomal RNAs, and 6 genes with unknown functions. In total, 50 repeat sequences were identified in *C. osmantha* cv 'yidan' cp genomes. Phylogenetic analysis showed that *C. osmantha* cv 'yidan' is more closely related to *Camellia vietnamensis* cv 'hongguo' and *Camellia oleifera* cv 'cenruan 3' than to *Camellia semiserrata* cv 'hongyu 1'. Our complete assembly of four *Camellia* cp genomes may contribute to breeding for high oil content and further biological discoveries. The results of this study provide a basis for the assembly of the entire chloroplast genome of *C. osmantha* cv 'yidan'.

1 Introduction

The genus *Camellia*, which is used worldwide as an ornamental plant and for tea, belongs to the family Theaceae (Vijayan et al. 2012, Yang et al. 2013, Hui et al. 2014). *Camellia* oil is less known worldwide despite its use in China as an edible oil, as well as in Japan. *Camellia* is one of the four main oil-bearing trees in the world, in addition to palm, olive, and coconut (Robards et al. 2009).

Through years of research and experimentation, the Guangxi Academy of Forestry discovered the new species *Camellia osmantha* (Ma et al. 2012). *C. osmantha* is easy to plant, grows rapidly, and has strong cold, heat, and drought tolerance (Ma et al. 2013, Liu et al. 2013) as well as high oil yield (Wang et al. 2014). *C. osmantha* cv 'yidan' is recognized as a new variety of *C. osmantha* (Ma 2020). The plant height and crown width of 6-year-old *C. osmantha* cv 'yidan' was 5.39 m and 7.17 m, respectively, and the oil production of a 5-year-old plant was 0.0590 kg·m⁻² (Liang et al. 2017), almost double the standard oil yield for *C. oleifera* cultivars (0.0325 kg·m⁻²). *Camellia* oil is also known as "eastern olive oil" because of the similarities in the chemical composition of *Camellia* and olive oils, with high amounts of oleic acid and linoleic acid, as well as low levels of saturated fats. At present, the total area of *C. osmantha* cv 'yidan' production is over 1500 ha, mainly in Qinzhou, Laibin, Yulin, Yunnan, and Hainan, China.

In China, the planting area of *Camellia oleifera* reaches 4,466,700 ha, and the oil production is 600,000 tons. *Camellia* oil production needs to be further developed, and improving plant photosynthetic characteristics is an effective way to increase plant yield. *C. osmantha* cv 'yidan' is a promising new species that produces 1590 kg of oil per hectare, doubling the standard oil productivity rate for *C. oleifera* cv 'cenruan 3' elite cultivars (750 kg·ha⁻¹) (Liang et al. 2017). Therefore, research on the photosynthetic characteristics based on the whole chloroplast genome sequence of *C. osmantha* cv 'yidan' is of great significance for improving production. At present, the chloroplast genome sequences of more than 20 plants in the genus *Camellia* have been published in NCBI, including species for ornamental purposes (Jun et al. 2013, Huang et al. 2014) and tea production and *Camellia oleifera*, which is mainly used to study the genetic evolutionary relationship of *Camellia* plants

The chloroplast (cp) genome is independent of the nuclear genome and exhibits maternal inheritance and semi-autonomous genetic characteristics (Guo et al. 2018). The structure of the cp genome in *Camellia* species is a typical four-segment, closed-loop structure, with a large single copy (LSC) region, a small single

copy (SSC) region, and two inverted repeats (IRs) of roughly the same length (Zheng et al. 2019). Among these structural regions, the IRs are the most stable, and the LSC has a higher mutation rate than the SSC. As the center of photosynthesis, the chloroplast genome is of great significance for revealing the mechanism and metabolic regulation of plant photosynthesis (Fang et al. 2010, Huang et al. 2013). Differences in plant photosystems can be used to improve the efficiency of light absorption and transformation and further increase plant yield (Zhang et al. 2011).

The coding regions of genes have a slower evolution rate, which is suitable for the analysis of relationships at the family and higher levels, while the non-coding regions have a faster mutation rate (Chen et al. 2018), which is more suitable for analyzing relationships at lower levels such as genera and species (Clegg et al. 1994, Zeng et al. 2017, Cui et al. 2019, Yang et al. 2019). Thus, the characteristics of the maternal and highly conserved genes of the chloroplast genome provide favorable conditions for studying the phylogeny of plants.

Research on the chloroplast genome of *Camellia* plants is currently limited to the use of some chloroplast genes for phylogenetic analysis. Here, we describe the whole chloroplast genome sequence of *C. osmantha* cv 'yidan' and three other *Camellia* species using the next-generation Illumina genome analyzer platform. The three representative species have notable phenotypic differences (including pericarp thickness, fruit size, seed yield, and oil content) and are widely cultivated in southern China. This study aimed to provide more information for the classification of *C. osmantha* cv 'yidan' by clarifying and comparing the cp genome sequences and structural variations between *C. osmantha* cv 'yidan' and three closely related *Camellia* species.

2 Materials and Methods

2.1 Sample Preparation, Sequencing and Chloroplast Genome Assembly

Fresh and healthy leaves of four *Camellia* species (*C. osmantha* cv 'yidan', *Camellia vietnamensis* cv 'hongguo', *Camellia oleifera* cv 'cenruan 3' and *Camellia semiserrata* cv 'hongyu 1') were sampled and used for complete cp genome sequencing. The four *Camellia* species were deposited in the *Camellia* oil Germplasm Resource (Latitude 22°55'51" Longitude 108°20'03"). A modified CTAB method was used to extract total genomic DNA from 50 mg of fresh leaves. To generate chloroplast assemblies, a 270-bp or 350-bp insertion library was constructed for each species, using TruSeq DNA sample preparation kits, and sequenced using Illumina technology with 150-bp paired-end sequencing mode at Kunming Institution of Botany, Chinese Academy of Sciences.

A total of 72 million raw reads were generated and made available in FASTQ format. The quality of the raw sequence reads was evaluated using the software package FastQC (Andrews 2010). The software Trimmomatic v0.36 was used for removal of adapter, contaminant, low-quality (Phred scores <30), and short (<36 bp) sequencing reads. The remaining high-quality sequencing reads were assembled de novo using the NOVOPlasty pipeline v2.7.2 with default parameters and based on a kmer size of 39 or 23 following the developer's suggestions, where the *psbA* gene of *C. oleifera* cv 'cenruan 3' was used as a seed input.

2.2 Chloroplast Genomic Annotation and Sequence Analyses

The assembled genomes of four species were originally annotated using PGA (Qu et al. 2019). The annotation results of codon positions and intron/exon boundaries were manually corrected by comparing with other known homologous genes (NC_023084.1) in the *Camellia* cp genome. The circular structures were mapped using the OGDRAW tool (Lohse et al. 2007). By aligning the IR/LSC and IR/SSC regions with homologous sequences from other *Camellia* species (NC_023084.1), their exact boundaries were determined.

Variation detection and evolutionary relationship analysis

Repeat structures including palindromic, forward, complement, and reverse repeats were searched with bibiserv software (<https://bibiserv.cebitec.uni-bielefeld.de/reputer>) with a repeat size of 15 bp and 90% or greater sequence identity. SSRs within the four cp genomes were detected using MISA software (<https://webblast.ipk-gatersleben.de/misa/index.php>). The following parameters were set in MISA: maximum length of sequence between two SSRs to register as compound SSR for 100 bp, with the parameters set at 10 for mononucleotides, 6 for dinucleotides, 5 for trinucleotides, and 5 for tetranucleotide, pentanucleotide, and hexanucleotide repeats.

We aligned the 114 *Camellia* and four other oil-producing species cp genome sequences using ClustalX. Unambiguously aligned DNA sequences were used for phylogenetic analyses, but ambiguously aligned regions were excluded. Maximum likelihood (ML) analyses were conducted using MEGA7. Bootstrap support (BS) values for individual clades were calculated by running 1,000 bootstrap replicates of the data. ML Heuristic Method searches were conducted with the Nearest-Neighbor-Interchange (NNI). The genetic relationship of the four *Camellia* cp genomes together with 108 available *Camellia* (Table 3) and four other oil-producing species cp genome sequences (GenBank accession no. JF937588.1 (*Ricinus communis* cultivar Hale), NC_016736.1 (*Ricinus communis*), GU931818.1 (*Olea europaea* cultivar Frantoio) and NC_013707.2) (*Olea europaea* cultivar Bianchera) were used to construct a maximum likelihood method (ML) tree by using MEGA 7 with default parameters (Tamura et al. 2011).

3 Results

3.1 The Structure of the Chloroplast Genomes of Four *Camellia* Species

The complete cp genomes of *C. semiserrata* cv'hongyu 1' (GenBank accession no. OP953553), *C. vietnamensis* cv'hongguo' (GenBank accession no. OP 953555), *C. osmantha* cv'yidan' (GenBank accession no. OP936137), and *C. oleifera* cv'cenruan 3' (GenBank accession no. OP953554) were sequenced using Illumina sequencing technology (Fig. 1). The cp genomes of the four species are composed of a circular DNA molecule ranging in size from 156,807 to 157,005 bp, with the typical quadripartite structure consisting of two inverted repeats (IRa and IRb) and LSC and SSC regions (Table 2).

The *C. semiserrata* cv'hongyu 1', *C. osmantha* cv'yidan', and *C. oleifera* cv'cenruan 3' cp genomes each contain 134 genes (81 protein-coding genes, 39 transfer RNA (tRNA) genes, and 8 ribosomal RNA (rRNA) genes, as well as 6 genes with unknown functions. The *C. vietnamensis* cv'hongguo' cp genome contains 136 genes (83 protein-coding genes, 39 tRNA genes, and 8 rRNA genes, as well as 6 genes with unknown functions, which includes two copies of the *rp12* gene. By contrast, *rp12* is not found in the other three species.

Among the 134 unique genes in *C. semiserrata cv'hongyu 1'*, *C. osmantha cv'yidan'*, and *C. oleifera cv'cenruan 3'*, 15 contain one intron (*petB*, *petD*, *atpF*, *ndhA*, *ndhB*, *rps12*, *rps16*, *rpl16*, *trnG-UCC*, *trnK-UUU*, *trnL-UAA*, *trnA-UGC*, *trnI-GAU*, *trnV-UAC*, and *rpoC1*), and 2 contain two introns (*clpP* and *ycf3*). Previous studies reported that *ycf3* is necessary for the stable accumulation of the photosystem I complex (Boudreau et al. 1997; Naver et al. 2001; Guo et al. 2018). Among the 135 unique genes in *C. vietnamensis cv'hongguo'*, 16 contain one intron (*petB*, *petD*, *atpF*, *ndhA*, *ndhB*, *rps12*, *rps16*, *rpl2*, *rpl16*, *trnG-UCC*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *trnA-UGC*, *trnI-GAU*, and *rpoC1*), and 2 contain two introns (*clpP* and *ycf3*). The gene maps of *C. osmantha cv'yidan'*, *C. semiserrata cv'hongyu 1'*, *C. oleifera cv'cenruan 3'*, and *C. vietnamensis cv'hongguo'* are shown in Fig. 1.

3.2 Expansion and Contraction of the Border Regions

The border regions and neighboring genes of the four *Camellia* cp genomes were compared to analyze the expansion and contraction of the connected regions (Fig. 2). The cp genomic structures, including gene type, gene order, and gene number, were conserved in *C. osmantha cv'yidan'* and *C. oleifera cv'cenruan 3'*, while the cp genomes of *C. vietnamensis cv'hongguo'* exhibited visible differences at the IRb/SSC/IRa/borders. The IRb region expanded into the gene *ycf1* with 1042–1068 bp in the IRb regions (1068 bp for *C. osmantha cv'yidan'* and *C. oleifera cv'cenruan 3'*, 1042 bp for *C. semiserrata cv'hongyu 1'*).

The IRa/SSC borders displayed large differences among the four cp genomes. The gene *ndhF* is located at the IRa/SSC or IRb/SSC junction, with 5–65 bp gaps between *ndhF* and the IR/SSC junction (5, 56, and 65 bp gaps in *C. semiserrata cv'hongyu 1'*, *C. osmantha cv'yidan'*, and *C. oleifera cv'cenruan 3'*, respectively). The *ndhF* and *ycf1* genes in *C. vietnamensis cv'hongguo'* are reversed in the IRb/SSC/IRa boundary region compared with the cp genome sequences of the other three species. *ndhF* in the SSC region was 56 bp from the IRb/LSC junction in *C. vietnamensis cv'hongguo'*. By contrast, the IRa/LSC and IRb/LSC boundary regions were relatively conserved in the four cp genomes. The gene *rpl2* formed another boundary by expanding into the IRa region in *C. vietnamensis cv'hongguo'*, leading to complete duplication of the gene within the IRs.

3.3 Long-Repeat and Simple Sequence Repeat (SSR) Analysis

We detected palindromic, forward, complementary, and reverse repeats in the four cp genomes. Overall, 50 repeat sequences were identified in all *Camellia* cp genomes, of which 23–24 palindromic repeats, 16–17 forward repeats, 7–9 reverse repeats, and 2–4 complementary repeats were separately found (Fig. S1(A)). The lengths of palindromic repeats ranged from 19 to 79 bp, the forward repeats ranged in length from 19–42 bp, the reverse repeats ranged in length from 19–23 bp, and the complementary repeats ranged in length from 19–20bp (Fig. S1(B–E))

In this study, we found 50, 51, 51, and 53 SSRs in the *C. semiserrata cv'hongyu 1'*, *C. osmantha cv'yidan'*, *C. vietnamensis cv'hongguo'*, and *C. oleifera cv'cenruan 3'* cp genomes, respectively (Fig. 3). These SSRs were mainly composed of adenine (A) or thymine (T) repeats and did not contain guanine (G) or cytosine (C) repeats. Moreover, the four cp genomes only contained mononucleotide repeats ranging from 10 to 17 bp.

3.4 Phylogenetic Analysis

We generated a phylogenetic tree using the nucleotide sequences of the cp genomes of 112 *Camellia* species and other oilseed crops using the maximum likelihood method (Fig. 4), and *Coffea arabica* (NC_008535.1)

was selected as an outgroup. *C. osmantha* cv'yidan' is most closely related to *C. vietnamensis* cv'hongguo' and *C. oleifera* cv'cenruan 3', which belong to the section *Oleifera* Chang.

4 Discussion

In this study, we sequenced the complete cp genomes of four *Camellia* species and annotated their sequences. Phylogenetic studies have shown that cp genome evolution includes nucleotide substitutions and structural changes (Feng et al. 2008, Haberle et al. 2008, Guo et al. 2018).

Some studies have shown that there are introns or gene deletions in the chloroplast genome (Downie et al. 1991, Downie et al. 1996, Graveley et al. 2001, Jansen et al. 2007, Ueda et al. 2007, Guisinger et al. 2010). Introns play an important role in the regulation of gene expression (Xu et al. 2017). They can increase gene expression levels in specific locations and at specific times (Le et al. 2003, Niu et al. 2011). The intron regulation mechanism has also been researched in other species (Callis et al. 1987, Emami et al. 2013). However, no studies have analyzed the association between intron loss and gene expression. The *chlB*, *chlL*, *chlN*, and *trnP-GGG* genes were missing in the four *Camellia* cp genomes but were found in several other angiosperm plastomes (Jansen et al. 2007, Green 2011, Mader et al. 2018). These four genes represent synapomorphies for flowering plants (Jansen et al. 2007). We found that *rpl2* was lost in the *C. semiserrata* cv'hongyu 1', *C. osmantha* cv'yidan', and *C. oleifera* cv'cenruan 3' cp genomes. Annotation of *rpl2* in *C. vietnamensis* cv'hongguo' showed that it encodes ribosomal protein L2. Whether *rpl2* can be used as a molecular marker for *C. vietnamensis* cv'hongguo' needs to be further verified with more *Camellia* species.

We found 15 genes that contained one intron and two genes that contained two introns (*ycf3* and *clpP*) in the *C. osmantha* cv'yidan' cp genomes. Introns can improve gene expression at specific times and specific locations (Le et al. 2003, Niu et al. 2011). The *ycf3* protein is necessary to stabilize the complex of photosystem I with the light-harvesting complex I (Boudreau et al. 1997, Naver et al. 2001). We therefore speculate that intron gain in *ycf3* may alter the expression of genes encoding the photosystem I assembly protein. In the next study, we will focus on comparing the photosynthesis-related genes of the four species. The *clpP* gene includes two introns. The intron gain in *clpP* may alter the regulation of genes encoding the *clp* protease proteolytic subunit. This phenomenon might be due to the increased evolutionary rates.

The *accD* gene encodes the heteroacetyl coenzyme A carboxylase (ACCase), a key enzyme involved in plant fatty acid biosynthesis (Kode et al. 2005, Nakkaew et al. 2008, Wicke et al. 2011, Zhang et al. 2016). Maliga (Maliga and Svab 2011) showed that *accD* in *Nicotiana glauca* was 1539 bp long. The *accD* sequence lengths were 1541, 1541, 1541, and 1532 bp in *C. oleifera* cv'cenruan 3', *C. semiserrata* cv'hongyu 1', *C. osmantha* cv'yidan', and *C. vietnamensis* cv'hongguo', respectively, suggesting that this gene has been conserved in plant cp genomes. Moreover, we observed no pseudogene formation of *accD* in the four *Camellia* cp genomes, consistent with the importance of fatty acid biosynthesis for these oil-producing plants.

The *rpoA* and *rpoC2* genes encode the alpha and beta subunits of plastid RNA polymerase (PEP), respectively, which is responsible for the transcription of most photosynthetic proteins. The light compensation point of *C. oleifera* cv'cenruan 3' is only $15.50 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$, and this species is more adapted to low light conditions compared to the other *Camellia* species (Ma et al. 2012); the light saturation point of *C. osmantha* cv'yidan' is

499.7 $\mu\text{mol} \cdot \text{m}^{-2}\text{s}^{-1}$, and this species is more adapted to high light conditions compared to the other *Camellia* species. These genes may play an important role in the evolution and differentiation of *Camellia* plants, particularly with regard to light requirements.

Phylogenetic relationships among four *Camellia* species revealed that *C. osmantha* cv 'yidan' is more closely related to *C. vietnamensis* cv 'hongguo' and *C. oleifera* cv 'cenruan 3' than to *C. semiserrata* cv 'hongyu 1', other *Camellia* species, and other oil crops. The results of this study provide an assembly of a whole chloroplast genome of *C. osmantha* cv 'yidan', which may be useful for future breeding and further biological discoveries. It will provide a theoretical basis for the improvement of *Camellia* oil yield and the determination of phylogenetic status.

Declarations

Disclosure statement

The authors are really grateful to the opened raw genome data from public database. The authors report no conflicts of interest and are responsible for the content and writing of the paper.

Ethics approval and consent to participate

Our study does not involve ethics approval and consent to participate.

Consent to publish

All authors read and approved the final manuscript.

Availability of data and materials

All the data involved in this article is true and reliable. The specimen of four *Camellia* species plants are deposited in the *Camellia oleifera* Germplasm Resource (<http://www.gxiky.com.cn/>, Bingqing Hao, bingqinghao_2021@caf.cn). The DNA samples are stored in the laboratory of Guangxi Academy of Forestry (<http://www.gxiky.com.cn/>, Bingqing Hao, bingqinghao_2021@caf.cn).

Competing interests

The authors declare that they have no competing interests.

Funding

Basic scientific research expenses of Guangxi Forestry Research Institute under Grant #Linke202302

Science Foundation of Guangxi Province, under Grant #2021GXNSFBA196085.

Data availability statement

The genome sequence data that support the findings of this study are openly available in the NCBI. *C. semiserrata* cv 'hongyu 1' (GenBank accession no. OP953553), *C. vietnamensis* cv 'hongguo' (GenBank

accession no. OP 953555), *C. osmantha* cv' *yidan* (GenBank accession no. OP936137), and *C. oleifera* cv' *cenruan 3* (GenBank accession no. OP953554).

Author Contributions

The research structure was designed by L.L., and H.Y.; B.H. prepared the sample and performed the experiments, analyzed the data and wrote the paper; L.L., G.C., and J.M. made revisions to the final manuscript. The final manuscript was read and corrected by all authors.

Acknowledgments

Thanks to Hong Yang for the analysis, assembling and annotation of sequencing sequences raw reads in the Kunming Institute of Botany, CAS. The sequencing of the four *Camellia* species involved in the article was performed on the Germplasm Bank of Wild Species.

References

1. Andrews S (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. Cambridge: Babraham Bioinformatics, Babraham Institute
2. Asaf S, Khan AL, Khan MA, et al (2017) Chloroplast genomes of *Arabidopsis halleri* ssp. *gemmifera* and *Arabidopsis lyrata* ssp. *petraea*: Structures and comparative analysis. *Nature* 7: 1-15
3. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinform* 30:2114-2120
4. Boudreau E, Takahashi Y, Lemieux C, et al. (1997) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *Embo J* 16: 6095-6104
5. Callis J, Fromm M, Walbot V (1987) Introns increase gene expression in cultured maize cells. *GenesDev* 1: 1183–1200
6. Clegg MT, Gaut BS, Learn GH, et al. (1994) Rates and patterns of chloroplast DNA evolution. *Proc. Natl Acad Sci U S A* 91: 6795-6801
7. Cui Y, Zhou J, Chen X, et al. (2019) Complete chloroplast genome and comparative analysis of three *Lycium* (Solanaceae) species with medicinal and edible properties. *Gene Reports* 17: 100464
8. Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18
9. Drummond AJ, Ashton B, Buxton S, et al. (2011) Geneious v5. 4. 2011. Geneious v5, 4
10. Downie SR, Llanas E, Katz-Downie DS (1996) Multiple independent losses of the *rpoC1* intron in angiosperm chloroplast DNAs. *Syst Bot* 21:135–151
11. Downie SR, Olmstead RG, Zurawski G, et al. (1991) Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: Molecular and phylogenetic implications. *Evolution* 45: 1245–1259
12. Emami S, Arumainayagam D, Korf I, Rose AB (2013) The effects of a stimulating intron on the expression of heterologous genes in *Arabidopsis thaliana*. *Plant biotechnology journal* 11: 555–563

13. Fang W, Yang JB, Yang SX, et al (2010) Phylogeny of *Camellia* sects. *Longipedicellata*, *Chrysantha* and *Longissima* (Theaceae) based on sequence data of four chloroplast DNA loci. *Acta Botanica Yunnanica* 32: 1-13
14. Feng Y, Cui L, Depamphilis CW, et al (2008) Gene rearrangement analysis and ancestral order inference from chloroplast genomes with inverted repeat. *BMC Genomics* 9, S25. 35
15. Graveley BR (2001) Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet* 17: 100–107
16. Green BR (2011) Chloroplast genomes of photosynthetic eukaryotes. *The plant journal* 66: 34-44
17. Guisinger MM, Chumley TW, Kuehl JV, Boore JL, Jansen RK (2010) Implications of the plastid genome sequence of *Typha* (Typhaceae, Poales) for understanding genome evolution in Poaceae. *J Mol Evol* 70: 149–166
18. Guo S, Guo L, Zhao W, et al. (2018) Complete chloroplast genome sequence and phylogenetic analysis of *Paeonia ostii*. *Molecules* 23: 246
19. Haberle RC, Fourcade HM, Boore JL, Jansen RK (2008) Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *J Mol Evol* 66: 350-361
20. Huang H, Tong Y, Zhang Q J, et al. (2013) Genome size variation among and within *Camellia* species by using flow cytometric analysis. *PLoS One* 8: e64981
21. Jansen RK, Cai Z, Raubeson LA, et al. (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad Sci U S A* 104: 19369-19374
22. Kode V, Mudd EA, lamtham S, Day A (2005) The tobacco plastid *accD* gene is essential and is required for leaf development. *Plant J* 44: 237-244
23. Kole C (2011) Wild crop relatives: Genomic and breeding resources cereals. New York: Springer Verlag. p. xxiii, 4973.
24. Le H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem.Sci* 28: 215-220
25. Liang HY, Hao BQ, Chen GC, et al. (2017) *Camellia* as an Oilseed Crop. *HortScience* 52: 488-497
26. Liu K, Zhou ZD, Wang DX, et al. (2013) Flooding Tolerance of Five *Camellia* Species. *Guangxi Forestry Research Science* 42: 329-332
27. Lohse M, DrechseL O, Kahlau S, Bock R (2013) Organellar Genome DRAW-A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* 41
28. Ma JL, Ye H, Ye CX (2012) A new species of *Camellia* sect. *Paracamellia*. *Guangxi Plant* 32:753-755
29. Ma JL, Zhang RQ, Ye H, He XY (2012) Photosynthetic characteristics among different *Camellia* species. *Nonwood Forest Research* 30:73-76+90
30. Ma JL, Zhang RQ, Ye H, He XY (2013) Semi-lethal temperature and cold tolerance & heat tolerance in *Camellia osmantha*. *Nonwood Forest Research* 31:150-152+175
31. Mader M, Pakull B, Blanc-Jolivet C, et al, (2018) Complete chloroplast genome sequences of four *Meliaceae* species and comparative analyses. *Int J Mol Sci* 19: 701

32. Maliga P, Svab Z (2011) Engineering the plastid genome of *Nicotiana glauca*, a diploid model species for plastid genetics. In *Plant Chromosome Engineering* (pp. 37-50). Humana Press, Totowa, NJ
33. Ming TL (1999). A systematic synopsis of the genus *Camellia*. *Acta Botanica Yunnanica* 21: 149-159
34. Ming TL (2000) Monograph of the genus *Camellia*. Kunming, China: Yunnan Science and Technology Press
35. Ming TL, Zhang WJ (1996) The evolution and distribution of genus *Camellia*. *Acta Bot Yunnan* 18: 1-13
36. Nakkaew A, Chotigeat W, Eksomtramee T, Phongdara A (2008) Cloning and expression of a plastid-encoded subunit, beta-carboxyltransferase gene (*accD*) and a nuclear-encoded subunit, biotin carboxylase of acetyl-CoA carboxylase from oil palm (*Elaeis guineensis* Jacq.). *Plant Sci* 175: 497-504
37. Naver H, Boudreau E, Rochaix JD (2001) Functional studies of *Ycf3*: its role in assembly of photosystem I and interactions with some of its subunits. *Plant Cell* 13: 2731-2745
38. Niu DK, Yang YF (2011) Why eukaryotic cells use introns to enhance gene expression: Splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol Direct* 6: 24
39. Qu XJ, Moore MJ, Li DZ, Yi TS (2019) PGA: a software package for rapid, accurate, and flexible batch annotation of plastomes. *Plant Methods* 15:50
40. Robards K, Prenzler P, Ryan D, Zhong H (2009) *Camellia* oil and tea oil, p. 313–343. In: R. Moreau and Kamal-Eldin (eds.). *Gourmet and health promoting specialty oils*. AOCS Press, Urbana, IL
41. Shi C, Liu Y, Huang H, Xia EH (2013) Contradiction between plastid gene transcription and function due to complex posttranscriptional splicing: an exemplary study of *ycf15* function and evolution in angiosperms. *PLoS One* 8, e59620.
42. Tamura K, Peterson D, Peterson N, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731-2739
43. Ueda M, Fujimoto M, Arimura SI (2007) Loss of the *rpl32* gene from the chloroplast genome and subsequent acquisition of a preexisting transit peptide within the nuclear gene in *Populus*. *Gene* 402: 51-56
44. Wang DX, Ye H, Ma JL, Zhou ZD (2014) Evaluation and selection of *Camellia osmantha* germplasm resources. *Nonwood Forest Research* 32: 159-162
45. Ma JL (2020) YiDan. Guangxi, Guangxi Forestry Research Institute, 2020-03-02
46. Wicke S, Schneeweiss GM, Depamphilis CW, et al. (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol Biol* 76: 273-297
47. Xu J, Chu Y, Liao B, Xiao S, Yin Q, Bai R, Su H, Dong L, Li X, Qian J, et al. (2017) *Panax ginseng* genome examination for ginsenoside biosynthesis. *Gigascience* 6: 1-15
48. Yang JB, Yang SX, Li HT, Yang J, Li DZ (2013) Comparative Chloroplast Genomes of *Camellia* Species. *PloS one*, 8, e73053
49. Yang Z, Huang Y, An, W, et al. (2019) Sequencing and Structural Analysis of the Complete Chloroplast Genome of the Medicinal Plant *Lycium chinense* Mill. *Plants* 8: 87
50. Zeng S, Zhou T, Han K, et al. (2017) The complete chloroplast genome sequences of six *Rehmannia* species. *Genes* 8: 103

51. Zhang JM, Liu J, Sun HL, Yu J, Wang JX, Zhou SL (2011) Nuclear and chloroplast SSR markers in *Paeonia delavayi* (Paeoniaceae) and cross-species amplification in *P. ludlowii*. *Am J Bot* 98:346-348
52. Zhang Y, Du L, Liu A, et al. (2016) The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front Plant Sci* 7: 306
53. Zhao P, Woeste KE (2011) DNA markers identify hybrids between butternut (*Juglans cinerea* L.) and Japanese walnut (*Juglans ailantifolia* Carr.). *Tree Genet Genomes* 7: 511-533
54. Zheng X, Ren C, Huang S et al. (2019) Structure and features of the complete chloroplast genome of *Melastoma dodecandrum*. *Physiology and molecular biology of plants: an international journal of functional plant biology* 25, 1043-1054

Tables

Table 1 Summary of *Camellia* chloroplast genome features

	<i>Camellia osmantha</i>	<i>Camellia vietnamensis</i>	<i>Camellia semiserrata</i>	<i>Camellia oleifera</i>
Genome size (bp)	156,981	157,003	156,807	157,005
LSC size (bp)	86,647	86,656	86,449	86,632
SSC size (bp)	18,284	18,297	18,256	18,291
IRa size (bp)	26,025	26,025	26,051	26,041
IRb size (bp)	26,025	26,025	26,051	26,041
Number of genes	134	136	134	134

Note: SSC(Small Single Copy Region); IRs(Inverted Repeats Region); LSC(Large Single Copy Region).

Table 2 List of genes in the three *Camellia* chloroplast genomes

Group of Genes		Gene Names	Number
Protein-coding genes	large subunit of Rubisco	<i>rbcL</i>	1
	photosystem1	<i>psaA, psaB, psaC, ycf1, psal</i>	5
	photosystem	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbl, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
	cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>	6
	ATP synthase	<i>atpA, atpB, atpE, atpF(*), atpH, atpI</i>	6
	NADH dehydrogenase	<i>ndhA*, ndhB(2)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12
	Envelope membrane protein	<i>cemA</i>	1
	ATP-dependent protease subunit P	<i>clpP**</i>	1
Ribosomal proteins	ribosomal small proteins	<i>rps2, rps3, rps4, rps7(2), rps8, rps11, rps12(3)*, rps14, rps15, rps16*, rps18, rps19</i>	15
	ribosomal large proteins	<i>rpl14, rpl16*, rpl20, rpl22, rpl23(2), rpl32, rpl33, rpl36</i>	9
RNA genes	tRNA genes	<i>trnA-UGC(2)*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC*, trnG-GCC, trnH-GUG, trnI-CAU(2), trnI-GAU(2)*, trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU(2), trnN-GUU(2), trnP-UGG, trnV-GAC(2), trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GGA, trnS-GCU, trnS-UGA, trnT-UGU, trnT-GGU, trnV-UAC*, trnV-GAC(2), trnW-CCA, trnY-GUA</i>	39
	rRNA genes	<i>rrn4.5(2), rrn5(2), rrn16(2), rrn23(2)</i>	8
Transcription/translation	Maturase	<i>matK</i>	1
	Subunit of acetyl-CoA carboxylase	<i>Accd</i>	1
	functions unknown (conserved open reading frames)	<i>ycf1, ycf2(2), ycf3**, ycf4, ycf15(2), ycf68</i>	8
	c-type cytochrome	<i>ccsA</i>	1

synthesis		
DNA-dependent RNA polymerase	<i>rpoA, rpoB, rpoC1*</i> , <i>rpoC2</i>	4
Translational initiation factor	<i>infA</i>	1
Total		134

* genes containing one intron; ** genes containing two introns; (2) genes present in two copies; (3) genes present in three copies.

rpl2: 2 copies in *C. vietnamensis* and 0 in the other three species.

Table 3 The list of accession number of the chloroplast genome sequences used in this study

Taxon	GenBank Accession Number	Taxon	GenBank Accession Number
<i>R.communis</i>	NC_016736.1	<i>C.chrysanthoides</i>	MW543443.1
<i>O.europaea</i>	NC_013707.2	<i>C.achrysantha</i>	MW543442.1
<i>R.communis</i>	JF937588.1	<i>C.brevistyla</i>	MW256435.1
<i>O.europaea</i>	GU931818.1	<i>C.pubipetala</i>	MW186719.1
<i>C.crapnelliana</i>	KF753632.1	<i>C.perpetua</i>	MW186718.1
<i>C.sinensis</i>	KF562708.1	<i>C.sinensis</i> var. <i>sinensis</i> cultivar <i>Tieguanyin</i>	MW148820.1
<i>C.taliensis</i> voucher HKAS:S.X.Yang3157	KF156839.1	<i>C.sinensis</i> isolate JM007 cultivar <i>Bantianyao</i>	MW046255.1
<i>C.yunnanensis</i> voucher HKAS:S.X.Yang1090	KF156838.1	<i>C.fascicularis</i>	MW026668.1
<i>C.pitardii</i> voucher HKAS:S.X.Yang3148	KF156837.1	<i>C.meiocarpa</i>	MT956593.1
<i>C.taliensis</i> voucher HKAS:S.X.Yang3158	KF156836.1	<i>C.sinensis</i> cultivar <i>Tieluohan</i>	MT773377.1
<i>C.impressinervis</i> voucher HKAS:S.X.Yang1080	KF156835.1	<i>C.sinensis</i> cultivar <i>Shuijingui</i>	MT773376.1
<i>C.danzaiensis</i> voucher HKAS:S.X.Yang3147	KF156834.1	<i>C.sinensis</i> cultivar <i>Rougui</i>	MT773375.1
<i>C.cuspidata</i> voucher HKAS:S.X.Yang3159	KF156833.1	<i>C.grandibracteata</i>	NC_024659.1
<i>C.sinensis</i>	KC143082.1	<i>C.crapnelliana</i>	NC_024541.1
<i>C.arabica</i>	NC_008535.1	<i>C.yunnanensis</i> voucherHKAS:S.X.Yang1090	NC_022463.1
<i>C.azalea</i>	KY856741.1	<i>C.pitardii</i> voucher HKAS:S.X.Yang3148	NC_022462.1
<i>C.luteoflora</i> voucher CLUTE20161220	KY626042.1	<i>C.impressinervis</i> voucherHKAS:S.X.Yang1080	NC_022461.1
<i>C.liberofilamenta</i> voucher CLIBE20161220	KY626041.1	<i>C.danzaiensis</i> voucherHKAS:S.X.Yang3147	NC_022460.1
<i>C.huana</i> voucher CHUAN20161220	KY626040.1	<i>C.cuspidata</i> voucher HKAS:S.X.Yang3159	NC_022459.1
<i>C.japonica</i>	KU951523.1	<i>C.taliensis</i> voucher HKAS:S.X.Yang3157	NC_022264.1
<i>C.sinensis</i> var. <i>sinensis</i>	KJ806281.1	<i>C.sinensis</i>	NC_020019.1
<i>C.sinensis</i> var. <i>pubilimba</i>	KJ806280.1	<i>C.japonica</i> strain <i>Huaheling</i>	MW602996.1

<i>C.sinensis</i> var. <i>dehungensis</i>	KJ806279.1	<i>C.debaoensis</i>	MW543445.1
<i>C.reticulata</i>	KJ806278.1	<i>C.pubipetala</i>	MW543444.1
<i>C.pubicosta</i>	KJ806277.1	<i>C.nitidissima</i>	NC_039645.1
<i>C.petelotii</i>	KJ806276.1	<i>C.gymnogyna</i>	NC_039626.1
<i>C.leptophylla</i>	KJ806275.1	<i>C.ptilophylla</i>	NC_038198.1
<i>C.grandibracteata</i>	KJ806274.1	<i>C.granthamiana</i>	NC_038181.1
<i>C.gymnogyna</i>	MH394406.1	<i>C.chekiangoleosa</i>	NC_037472.1
<i>C.gymnogyna</i>	MH394405.1	<i>C.japonica</i> strain S288C	NC_036830.1
<i>C.gymnogyna</i>	MH394404.1	<i>C.azalea</i>	NC_035574.1
<i>C.gymnogyna</i>	MH394403.1	<i>C.reticulata</i>	NC_024663.1
<i>C.nitidissima</i>	MH382827.1	<i>C.pubicosta</i>	NC_024662.1
<i>C.renshanxiangiae</i>	MH253889.1	<i>C.petelotii</i>	NC_024661.1
<i>C.sinensis</i>	MH042531.1	<i>C.leptophylla</i>	NC_024660.1
<i>C.ptilophylla</i>	MG797642.1	<i>C.kissii</i>	NC_053915.1
<i>C.granthamiana</i>	MG782842.1	<i>C.fascicularis</i>	NC_053896.1
<i>C.chekiangoleosa</i>	MG431968.1	<i>C.yuhsienensis</i>	NC_053622.1
<i>C.japonica</i> strain S288C	MF850254.1	<i>C.gauchowensis</i>	NC_053541.1
<i>C.oleifera</i>	MF541730.2	<i>C.brevistyla</i>	NC_052752.1
<i>C.sinensis</i> sangmok	LC488797.1	<i>C.amplexicaulis</i>	NC_051559.1
<i>C.grandibracteata</i>	KJ806274.1	<i>C.rhytidophylla</i>	NC_050389.1
<i>C.lungzhouensis</i>	MN579509.2	<i>C.fraterna</i>	NC_050388.1
<i>C.tachangensis</i> cultivar Xingyi6	MN327576.1	<i>C.anlungensis</i> voucher CANLU20191106	NC_050354.1
<i>C.sinensis</i> cultivar Baiye 1	MN086819.1	<i>C.renshanxiangiae</i>	NC_041672.1
<i>C.weiningensis</i> voucher CwCPF1-201901	MK820035.1	<i>C.sasanqua</i>	NC_041473.1
<i>C.japonica</i> isolate Jeju Island	MK353211.1	<i>C.sinensis</i> var. <i>assamica</i>	MH394407.1
<i>C.japonica</i> isolate Soyeonpyeongdo	MK353210.1	<i>C.sinensis</i> cultivar Dahongpao	MT773374.1
<i>C.sasanqua</i>	MH782189.1	<i>C.sinensis</i> cultivar Baijiguan	MT773373.1
<i>C.sinensis</i>	MH460639.1	<i>C.yuhsienensis</i>	MT665973.1
<i>C.sinensis</i> var. <i>assamica</i>	MH394410.1	<i>C.rhytidophylla</i>	MT663343.1

<i>C.sinensis</i> var. <i>assamica</i>	MH394409.1	<i>C.fraterna</i>	MT663342.1
<i>C.sinensis</i> var. <i>assamica</i>	MH394408.1	<i>C.chuongtsoensis</i>	MT663341.1
<i>C.amplexicaulis</i>	MT317095.1	<i>C.sinensis</i> cultivar <i>Wuyi Narcissus</i>	MT612435.1
<i>C.anlungensis</i> voucher CANLU20191106	MN756594.1	<i>C.gauchowensis</i>	MT449927.1
<i>C.brevistyla</i>	MN640791.1	<i>C.kissii</i>	MN635793.1

Figures

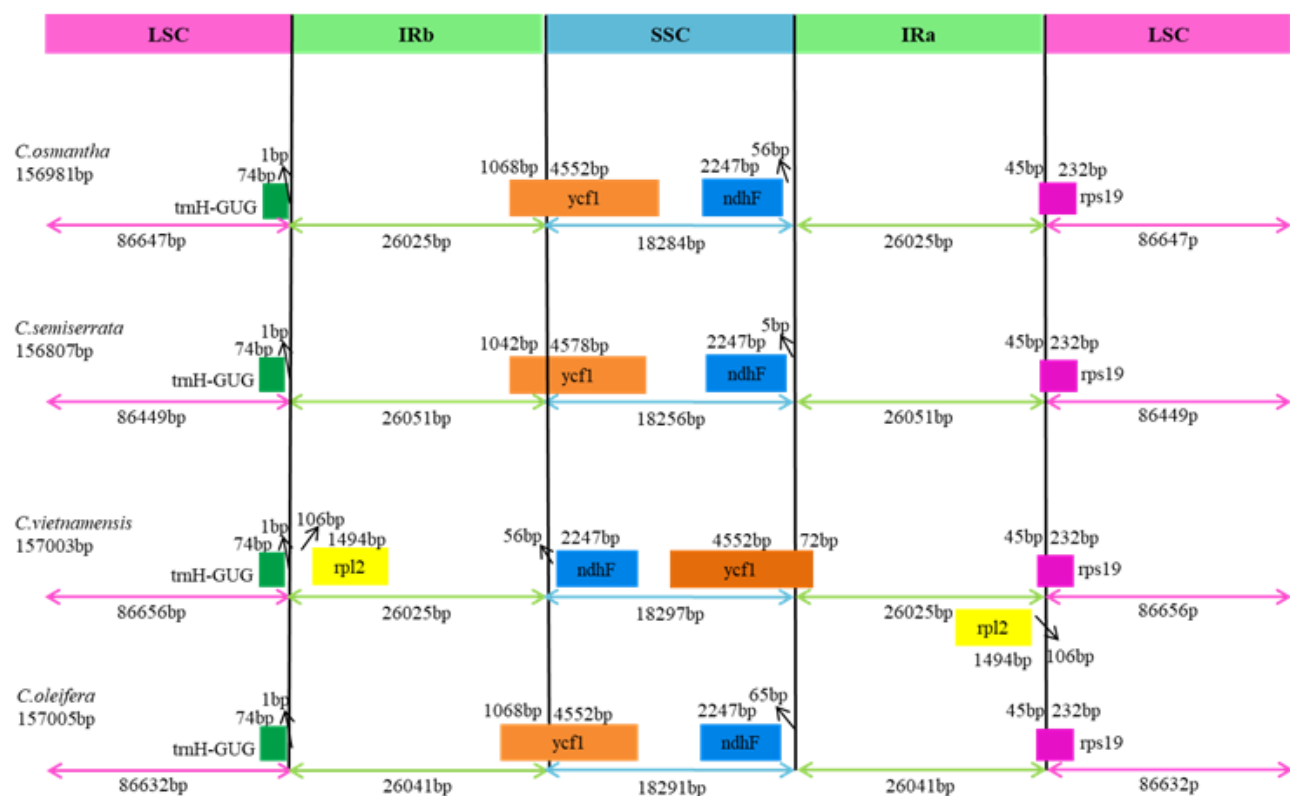
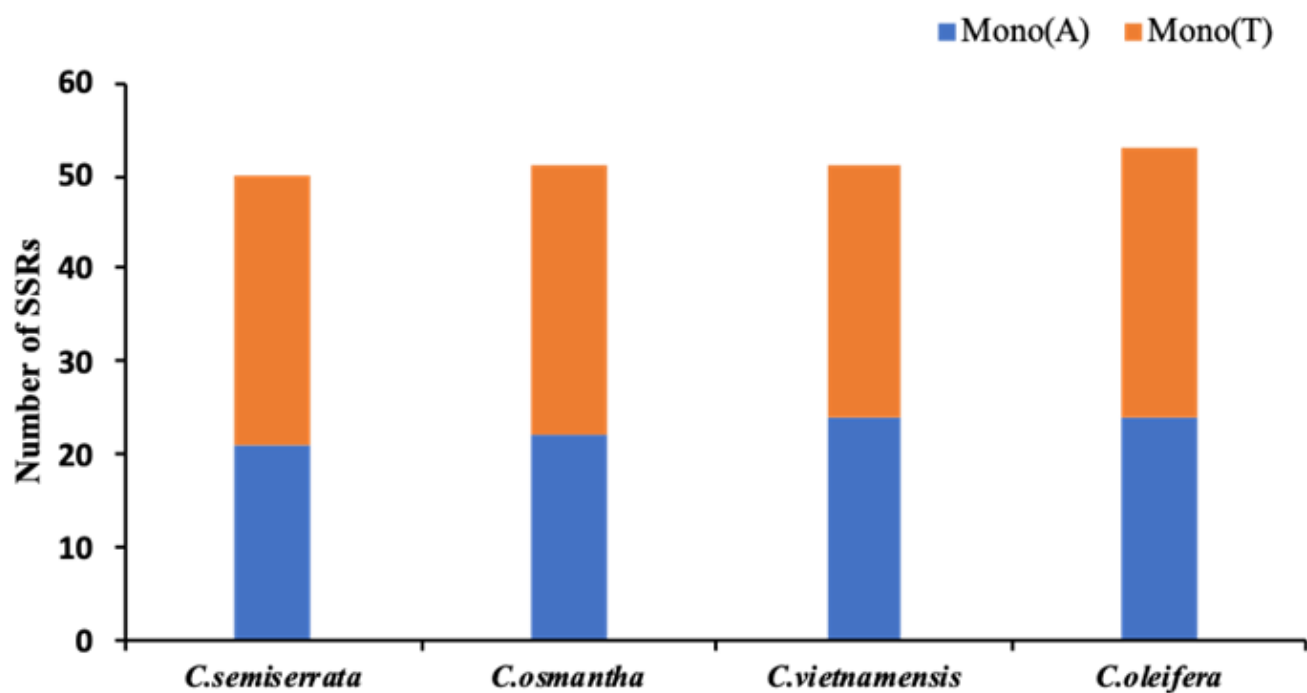


Figure 2

Comparison of the SSC, IRs, and LSC border regions among the four *Camellia* cp genomes

Note: SSC(Small Single Copy Region); IRs(Inverted Repeats Region); LSC(Large Single Copy Region).



Number of SSR motifs in different *Camellia* cp genomes

Phylogenetic tree of *Camellia* and other related oilseed species by using the maximum likelihood method

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalMaterials.docx](#)