

1 **Genome assembly of the rare and endangered Grantham's camellia, *Camellia***  
2 ***granthamiana***

3 Hong Kong Biodiversity Genomics Consortium

4 **Project Coordinator and Co-Principal Investigators:** Jerome H.L. Hui<sup>1</sup>, Ting Fung Chan<sup>2</sup>,  
5 Leo L. Chan<sup>3</sup>, Siu Gin Cheung<sup>4</sup>, Chi Chiu Cheang<sup>5,6</sup>, James K.H. Fang<sup>7</sup>, Juan Diego  
6 Gaitan-Espitia<sup>8</sup>, Stanley C.K. Lau<sup>9</sup>, Yik Hei Sung<sup>10,11</sup>, Chris K.C. Wong<sup>12</sup>, Kevin Y.L. Yip<sup>13,14</sup>,  
7 Yingying Wei<sup>15</sup>

8 **DNA extraction, library preparation and sequencing:** Sean T.S. Law, Wai Lok So<sup>1</sup>

9 **Genome assembly and gene model prediction:** Wenyan Nong<sup>1</sup>

10 **Genome analysis:** Sean T.S. Law<sup>1</sup>, Wenyan Nong<sup>1</sup>

11 **Sample collector, animal culture and logistics:** David T.W. Lau<sup>16</sup>, Ho Yin Yip<sup>1</sup>

12 1. School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory  
13 of Agrobiotechnology, Institute of Environment, Energy and Sustainability, The Chinese  
14 University of Hong Kong, Hong Kong, China

15 2. School of Life Sciences, State Key Laboratory of Agrobiotechnology, The Chinese  
16 University of Hong Kong, Hong Kong SAR, China

17 3. State Key Laboratory of Marine Pollution and Department of Biomedical Sciences, City  
18 University of Hong Kong, Hong Kong SAR, China

19 4. State Key Laboratory of Marine Pollution and Department of Chemistry, City University of  
20 Hong Kong, Hong Kong SAR, China

21 5. Department of Science and Environmental Studies, The Education University of Hong  
22 Kong, Hong Kong SAR, China

23 6. EcoEdu PEI, Charlottetown, PE, C1A 4B7, Canada

24 7. Department of Food Science and Nutrition, Research Institute for Future Food, and State  
25 Key Laboratory of Marine Pollution, The Hong Kong Polytechnic University, Hong Kong  
26 SAR, China

27 8. The Swire Institute of Marine Science and School of Biological Sciences, The University  
28 of Hong Kong, Hong Kong SAR, China

29 9. Department of Ocean Science, The Hong Kong University of Science and Technology,  
30 Hong Kong SAR, China

31 10. Science Unit, Lingnan University, Hong Kong SAR, China

32 11. School of Allied Health Sciences, University of Suffolk, Ipswich, IP4 1QJ, UK

33 12. Croucher Institute for Environmental Sciences, and Department of Biology, Hong Kong  
34 Baptist University, Hong Kong SAR, China

35 13. Department of Computer Science and Engineering, The Chinese University of Hong  
36 Kong, Hong Kong SAR, China

37 14. Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA.

38 15. Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China

39 16. Shiu-Ying Hu Herbarium, School of Life Sciences, The Chinese University of Hong  
40 Kong, Hong Kong SAR, China

41

42 Correspondence on behalf of the consortium: jeromehui@cuhk.edu.hk

43

44

45

46

47

48

## 49 **Abstract**

50 The Grantham's camellia (*Camellia granthamiana* Sealy) is a rare and endangered tea  
51 species that is endemic to southern China, and was first discovered in Hong Kong in 1955.  
52 Despite its high conservation value, genomic resources of *C. granthamiana* remain limited.  
53 Here, we present a chromosome-scale draft genome of the tetraploid *C. granthamiana* ( $2n =$   
54  $4x = 60$ ) using a combination of PacBio long read sequencing and Omni-C data. The  
55 assembled genome size is  $\sim 2.4$  Gb with most sequences anchored to 15 pseudochromosomes  
56 that resemble a monoploid genome. The genome is of high contiguity, with a scaffold N50 of  
57 139.7 Mb, and high completeness with a 97.8% BUSCO score. Gene model prediction  
58 resulted in a total 76,992 protein-coding genes with a BUSCO score of 85.9%. 1.65 Gb of  
59 repeat content was annotated, which accounts for 68.48% of the genome. The Grantham's  
60 camellia genome assembly provides a valuable resource for future investigations on its  
61 biology, ecology, phylogenomic relationships with other *Camellia* species, as well as set up a  
62 foundation for further conservation measures.

63

## 64 **Introduction**

65 *Camellia* is a large genus in the family Theaceae with more than 230 described  
66 species (POWO, 2021). While some camellias are well-known for their ornamental and  
67 economical values as tea and woody-oil producing plants that derived into tens of thousands  
68 of cultivars (Wang et al., 2021), more than 60 *Camellia* species were regarded as globally  
69 threatened due to natural habitat fragmentation or loss and small population size (Beech et al.,  
70 2017). The Grantham's camellia (*Camellia granthamiana*) (Figure 1A) is a rare species once  
71 discovered in Hong Kong and named after Sir Alexander Grantham and is narrowly  
72 distributed in Hong Kong and Guangdong, China (Beech et al., 2017). It is listed as  
73 vulnerable in the IUCN Red List and recorded as endangered in the China Plant Red Data  
74 Book (Fu & Chin, 1992). In Hong Kong, the Grantham's camellia is a protected species by  
75 law and has been actively being propagated and reintroduced to the wild by the Agriculture,  
76 Fisheries and Conservation Department (Hu, 2003).

77

## 78 **Context**

79 In view of the high conservation value of Grantham's camellia, several molecular  
80 studies have been previously conducted. They include sequencing the chloroplast genomes of  
81 *C. granthamiana* (Jiang et al., 2019; Li et al., 2018), using pan-transcriptomes to reconstruct  
82 the phylogeny of over a hundred of *Camellia* species (Wu et al., 2022), and population  
83 genetics study (Chen et al., 2023). However, nuclear genomic resources of *C. granthamiana*  
84 remain lacked. While most *Camellia* species possess a karyotype of  $2n = 30$ , *C.*  
85 *granthamiana* is one of the exceptions with a karyotype of  $2n = 4x = 60$  (Huang et al., 2013;

86 Kondo et al., 1977).

87 In Hong Kong, *C. granthamiana* was chosen as one of the listed species for  
88 sequencing in the Hong Kong Biodiversity Genomics Consortium (a.k.a. EarthBioGenome  
89 Project Hong Kong), which is formed by investigators from eight publicly funded universities.  
90 Herein, we report the genome assembly of *C. granthamiana* which can serve as a solid  
91 foundation for further investigations of this rare and endangered species.

92

### 93 **Methods**

#### 94 ***Sample collection and high molecular weight DNA extraction***

95 Fresh leaf tissues were sampled in transplanted individual on the campus of the  
96 Chinese University of Hong Kong. High molecular weight (HMW) genomic DNA was  
97 isolated from 1g leaf tissues using a CTAB pretreatment followed by NucleoBond HMW  
98 DNA kit (Macherey Nagel Item No. 740160.20). Briefly, the tissues were ground with liquid  
99 nitrogen and digested in 5 mL CTAB buffer (Doyle & Doyle, 1987) with an addition of 1%  
100 polyvinylpyrrolidone (PVP) for 1 h. The lysate was treated with RNase A, followed by an  
101 addition of 1.6 mL of 3M potassium acetate and two round of chloroform:IAA (24:1) washes.  
102 The supernatant was transferred to a new 50 mL tube using a wide-bore tip. H1 buffer from  
103 the NucleoBond HMW DNA kit was added to the supernatant for a total volume of 6 mL  
104 mixture, from which the DNA was isolated by following the manufacturer's protocol. After  
105 the DNA was eluted with 60  $\mu$ L elution buffer (PacBio Ref. No. 101-633-500), quality check  
106 was carried out with NanoDrop<sup>TM</sup> One/OneC Microvolume UV-Vis Spectrophotometer,  
107 Qubit<sup>®</sup> Fluorometer, and overnight pulse-field gel electrophoresis.

108

#### 109 ***Pacbio library preparation and sequencing***

110 The qualified DNA was sheared with a g-tube (Covaris Part No. 520079) for 6 passes  
111 of centrifugation at 1,990  $\times$  g for 2 min and was subsequently purified with SMRTbell<sup>®</sup>  
112 cleanup beads (PacBio Ref. No. 102158-300). 2  $\mu$ L sheared DNA was taken for fragment size  
113 examination through overnight pulse-field gel electrophoresis. Two SMRTbell libraries were  
114 constructed with the SMRTbell<sup>®</sup> prep kit 3.0 (PacBio Ref. No. 102-141-700), following the  
115 manufacturer's protocol. The final library was prepared with the Sequel<sup>®</sup> II binding kit 3.2  
116 (PacBio Ref. No. 102-194-100) and was loaded with the diffusion loading mode with the  
117 on-plate concentration set at 90 pM on the Pacific Biosciences SEQUEL IIe System, running  
118 for 30-hour movies to output HiFi reads. In total, three SMRT cells were used for the  
119 sequencing. Details of the resulting sequencing data are summarized in Supplementary  
120 Information 1.

121

#### 122 ***Omni-C library preparation and sequencing***

123 Nuclei was isolated from 3 g fresh leaf tissues ground with liquid nitrogen using the

124 PacBio protocol modified from Workman et al. (2018)  
125 (<https://www.pacb.com/wp-content/uploads/Procedure-checklist-Isolating-nuclei-from-plant-tissue-using-TissueRuptor-disruption.pdf> ). The nuclei pellet was snap-frozen with liquid  
126 nitrogen and stored at -80 °C. Upon Omni-C library construction, the nuclei pellet was  
127 resuspended in 4 mL 1X PBS buffer and processed with the Dovetail® Omni-C® Library  
128 Preparation Kit (Dovetail Cat. No. 21005) by following the manufacturer's procedures. The  
129 concentration and fragment size of the resulting library was assessed by Qubit® Fluorometer  
130 and TapeStation D5000 HS ScreenTape, respectively. The qualified library was sent to  
131 Novogene and sequenced on an Illumina HiSeq-PE150 platform. Details of the resulting  
132 sequencing data are summarized in Supplementary Information 1.  
133

134

### 135 ***Genome assembly and gene model prediction***

136 *De novo* genome assembly was first proceeded with Hifiasm (Cheng et al., 2021) and  
137 then was processed with searching against the NT database with BLAST to remove possible  
138 contaminations using BlobTools (v1.1.1) (Laetsch & Blaxter, 2017). Subsequently, haplotypic  
139 duplications were removed according to the depth of HiFi reads using “purge\_dups” (Guan et  
140 al., 2020). Proximity ligation data from Omni-C were used to scaffold the assembly with  
141 YaHS (Zhou et al., 2022).

142 Gene models were trained, predicted and updated by funannotate (Palmer & Stajich,  
143 2020) with the following parameters “--repeats2evm --protein\_evidence uniprot\_sprot.fasta  
144 --genemark\_mode ET --optimize\_augustus --organism other --max\_intronlen 350000”.  
145 Seven RNA sequencing data were downloaded from NCBI (SRA Accessions: SRR16685015,  
146 SRR16685016, SRR16685017, SRR19086193, SRR19266768, SRR24821546, and  
147 SRR24821547) and aligned to the repeat soft-masked genome using Hisat2 to run the  
148 genome-guided Trinity (Grabherr et al., 2011), from which 289,554 transcripts were derived.  
149 The Trinity transcript alignments were converted to GFF3 format and used as input to run the  
150 PASA alignment to generate PASA models trained by TransDecoder, which were screened  
151 using Kallisto TPM data. The PASA gene models were used to train Augustus in the  
152 funannotate-predict step. The predicted gene models were combined from various prediction  
153 sources, including GeneMark, high-quality Augustus predictions (HiQ), pasa, Augustus,  
154 GlimmerHM and snap, and were integrated to produce the annotation files with Evidence  
155 Modeler. UTRs were further captured in the funannotate-update step using PASA.  
156

157

### ***Repeat annotation***

158 The annotation of transposable elements (TEs) were performed by the Earl Grey TE  
159 annotation pipeline (version 1.2, <https://github.com/TobyBaril/EarlGrey>) (Baril et al., 2022).  
160

### 161 ***Macrosynteny analysis***

162 The longest gene transcripts from the predicted gene models of *Camellia*  
163 *granthamiana* and *Camellia sinensis* (accession number: GWHASIV00000000; Zhang et al.,  
164 2021) were used to retrieve orthologous gene pair with reciprocal BLASTp (e-value 1e-5)  
165 using diamond (v2.0.13) (Buchfink et al., 2021). The BLAST output was passed to MCScanX  
166 (Wang et al., 2012) to infer macrosynteny of the pseudochromosomes between *C.*  
167 *granthamiana* and *C. sinensis* with default parameters.

168

## 169 **Results and discussion**

### 170 *Genome assembly of Camellia granthamiana*

171 A total of 54.4 Gb HiFi reads was yielded from PacBio sequencing with an average  
172 length of 10,731 bp (Table 1; Supplementary Information 1). Together with 233.8 Gb  
173 Omni-C data, the genome of *Camellia granthamiana* was assembled with a final size of  
174 2,412.5 Mb, from which 79.87% of the sequences were anchored into 15  
175 pseudochromosomes (Figure 1B-1D; Supplementary Information 1). The scaffold N50 was  
176 139.7 Mb and the BUSCO score was 97.8% (Figure 1B; Table 1). Gene model prediction  
177 yielded a total of 76,992 protein-coding genes with a mean length of 301 bp and BUSCO  
178 score of 85.9%.

179 Repeat content analysis annotated 1.65 Gb of transposable elements (TEs),  
180 comprising 68.48% of the *C. granthamiana* genome. Among the classified TEs, LTR  
181 retransposons accounted for the largest proportion (20.99%), followed by DNA transposons  
182 (5.30%), LINE (1.60%) and Rolling-circle transposons (1.21%) (Figure 1D; Table 2). The  
183 large proportion of repeat content in the *C. granthamiana* genome is comparable to other tea  
184 species, such as the Tieguanyin cultivar of *Camellia sinensis* (78.2%) (Zhang et al., 2021),  
185 wild oil-Camellia *Camellia oleifera* (76.1%) (Lin et al., 2022), and *Camellia chekiangoleosa*  
186 (79.09%) (Shen et al., 2022).

187

### 188 *Macrosynteny between Camellia granthamiana and Camellia sinensis*

189 Macrosynteny analysis revealed a 1-to-1 pair relationship between the 15  
190 pseudochromosomes of *C. granthamiana* and that of *C. sinensis* (Figure 2). This indicates that  
191 the assembled 15 pseudochromosomes resemble a monoploid genome of the tetraploid *C.*  
192 *granthamiana*.

193

### 194 **Conclusion and future perspective**

195 This study presents the first *de novo* genome assembly of the rare and endangered *C.*  
196 *granthamiana*. This valuable genome resource is of great potential for the use in future  
197 studies on the conservation biology of the Grantham's camellia, its relationship with other  
198 *Camellia* species from a phylogenomic perspective and further investigations on the  
199 biosynthesis of secondary metabolites of tea species.

200

201 **Data validation and quality control**

202 For HMW DNA and Pacbio library samples, NanoDrop™ One/OneC Microvolume  
203 UV-Vis Spectrophotometer, Qubit® Fluorometer, and overnight pulse-field gel electrophoresis  
204 were used for quality control. The quality of Omni-C library was checked by Qubit®  
205 Fluorometer and TapeStation D5000 HS ScreenTape.

206 During genome assembly, BlobTools (v1.1.1) (Laetsch & Blaxter, 2017) was  
207 employed to remove possible contaminations (Supplementary Information 2). The resulting  
208 genome assembly was run with Benchmarking Universal Single-Copy Orthologs (BUSCO,  
209 v5.5.0) (Manni et al., 2021) with the Viridiplantae dataset (Viridiplantae Odb10) to assess the  
210 completeness of the genome assembly and gene annotation.

211

212 **Disclaimer**

213 The genomic data generated in this study was not fully haplotype-resolved for a  
214 tetraploid genome and the genome heterozygosity was not assessed.

215

216 **Data availability**

217 The final genome assembly in this study was submitted to NCBI under accession  
218 number JAXFYN000000000. The raw reads generated were deposited in the NCBI database  
219 under the SRA accessions SRR26895683 and SRR26909376. The genome annotation  
220 files were uploaded to Figshare (<https://figshare.com/s/4b13376ad27ae0647fd1>).

221

222 **Authors' contribution**

223 JHLH, TFC, LLC, SGC, CCC, JKHF, JDG, SCKL, YHS, CKCW, KYLY and YW  
224 conceived and supervised the study. DTWL collected the sample materials; STSL and WLS  
225 performed DNA extraction, library preparation and genome sequencing; HYY facilitated the  
226 logistics of samples; WN performed genome assembly, gene model prediction and genome  
227 quality check analyses; STSL carried out macrosynteny analysis.

228

229 **Competing interest**

230 The authors declare that they do not have competing interests.

231

232 **Funding**

233 This work was funded and supported by the Hong Kong Research Grant Council  
234 Collaborative Research Fund (C4015-20EF), CUHK Strategic Seed Funding for  
235 Collaborative Research Scheme (3133356) and CUHK Group Research Scheme (3110154).

236

237 **References**

238 1. Baril T, Imrie RM, Hayward A. Earl Grey: a fully automated user-friendly transposable  
239 element annotation and analysis pipeline. bioRxiv. 2022.  
240 <https://doi.org/10.1101/2022.06.30.498289>

241 2. Beech E, Barstow M, Rivers M. The red list of Theaceae. Botanic Gardens Conservation  
242 International; 2017.

243 3. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using  
244 DIAMOND. Nature methods. 2021;18(4):366-8.

245 4. Chen S, Li W, Li W, Liu Z, Shi X, Zou Y, Liao W, Fan Q. Population genetics of *Camellia*  
246 *granthamiana*, an endangered plant species with extremely small populations in China.  
247 Frontiers in Genetics. 2023;14.

248 5. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly  
249 using phased assembly graphs with hifiasm. Nature methods. 2021;18(2):170-5.

250 6. Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf  
251 tissue. Phytochemical bulletin. 1987.

252 7. Fu L, Chin CM. China plant red data book. Science press; 1992.

253 8. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing  
254 haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36(9):2896-8.

255 9. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L,  
256 Raychowdhury R, Zeng Q, Chen Z. Full-length transcriptome assembly from RNA-Seq  
257 data without a reference genome. Nature biotechnology. 2011;29(7):644-52.

258 10. Hu Q. Rare and Precious Plants of Hong Kong. Agriculture Fisheries and Conservation  
259 Department, the Government of the Hong Kong Special Administrative Region; 2003.

260 11. Huang H, Tong Y, Zhang QJ, Gao LZ. Genome size variation among and within *Camellia*  
261 species by using flow cytometric analysis. PLoS One. 2013;8(5):e64981.

262 12. Jiang Z, Jiao P, Qi Z, Qu J, Guan S. The complete chloroplast genome sequence of  
263 *Camellia granthamiana*. Mitochondrial DNA Part B. 2019;4(2):4113-5.

264 13. Kondo K. Chromosome numbers in the genus *Camellia*. Biotropica. 1977:86-94.

265 14. Kong W, Wang Y, Zhang S, Yu J, Zhang X. Recent Advances in assembly of plant  
266 complex genomes. Genomics, Proteomics & Bioinformatics. 2023;21(3):427-439.

267 15. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies.  
268 F1000Research. 2017;6(1287):1287.

269 16. Li W, Shi X, Guo W, Banerjee AK, Zhang Q, Huang Y. Characterization of the complete  
270 chloroplast genome of *Camellia granthamiana* (Theaceae), a vulnerable species endemic  
271 to China. Mitochondrial DNA Part B. 2018;3(2):1139-40.

272 17. Lin P, Wang K, Wang Y, Hu Z, Yan C, Huang H, Ma X, Cao Y, Long W, Liu W, Li X. The  
273 genome of oil-Camellia and population genomics analysis provide insights into seed oil  
274 domestication. Genome Biology. 2022;23:1-21.

275

276 18. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and  
277 streamlined workflows along with broader and deeper phylogenetic coverage for scoring  
278 of eukaryotic, prokaryotic, and viral genomes. Molecular biology and evolution.  
279 2021;38(10):4647-54.

280 19. Palmer JM, Stajich J. Funannotate v1. 8.1: Eukaryotic genome annotation. Zenodo  
281 <https://doi.org/10.5281/zenodo.2020;4054262>.

282 20. POWO. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew.  
283 Published on the Internet. 2024. <http://www.plantsoftheworldonline.org/>. Accessed 2 Jan  
284 2024.

285 21. Rivers MC, & Wheeler L. 2015. *Camellia granthamiana*. The IUCN Red List of  
286 Threatened Species 2015: e.T62053240A62053244.  
287 <https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T62053240A62053244.en>. Accessed  
288 on 13 December 2023.

289 22. Shen TF, Huang B, Xu M, Zhou PY, Ni ZX, Gong C, Wen Q, Cao FL, Xu LA. The reference genome of *Camellia chekiangoleosa* provides insights into *Camellia* evolution  
290 and tea oil biosynthesis. Horticulture Research. 2022;9:uhab083.

291 23. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H,  
292 Kissinger JC. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny  
293 and collinearity. Nucleic acids research. 2012;40(7):e49.

294 24. Wang Y, Zhuang H, Shen Y, Wang Y, Wang Z. The dataset of *Camellia* cultivars names in  
295 the world. Biodiversity Data Journal. 2021;9.

296 25. Workman R, Timp W, Fedak R, Kilburn D, Hao S, Liu K. High molecular weight DNA  
297 extraction from recalcitrant plant species for third generation sequencing. *Protocol  
298 Exchange* (published online 18 June 2018).  
300 <https://protocolexchange.researchsquare.com/article/nprot-6785/v1>.

301 26. Wu Q, Tong W, Zhao H, Ge R, Li R, Huang J, Li F, Wang Y, Mallano AI, Deng W, Wang  
302 W. Comparative transcriptomic analysis unveils the deep phylogeny and secondary  
303 metabolite evolution of 116 *Camellia* plants. The Plant Journal. 2022;111(2):406-21.

304 27. Zhang X, Chen S, Shi L, Gong D, Zhang S, Zhao Q, Zhan D, Vasseur L, Wang Y, Yu J,  
305 Liao Z. Haplotype-resolved genome assembly provides insights into evolutionary history  
306 of the tea plant *Camellia sinensis*. Nature Genetics. 2021;53(8):1250-9.

307 28. Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool.  
308 Bioinformatics. 2023;39(1):btac808.

309

310 **Table 1.** Genome statistics and sequencing information.

311 **Table 2.** Summary of classified transposable elements in the genome.

312

313 **Figure 1.** Genomic information of *Camellia granthamiana*. **A)** Picture of *Camellia*

314 *granthamiana*; **B)** Summary of genome statistics; **C)** Omni-C contact map of the genome  
315 assembly; **D)** Information of 15 pseudochromosomes; **E)** Pie chart (Top) and repeat  
316 landscape plot (bottom) of repetitive elements in the genome.

317

318 **Figure 2.** Macrosynteny dot plot between *Camellia granthamiana* and *Camellia sinensis*.

319

320 **Supplementary Information 1.** Summary of genomic sequencing data.

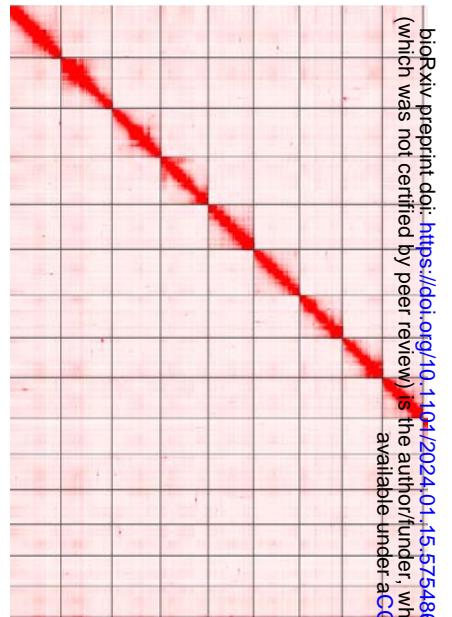
321 **Supplementary Information 2.** Genome assembly QC and contaminant/cobiont detection.



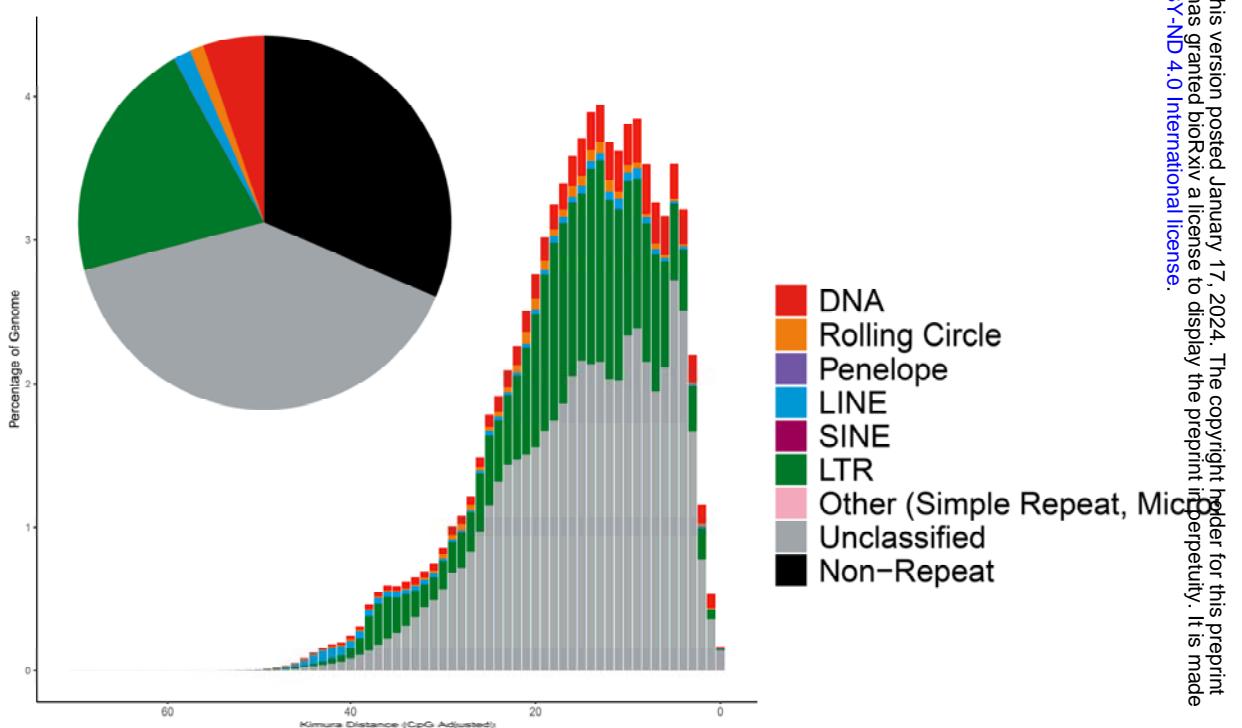
B)

	<i>Camellia granthamiana</i>
Genome size (bp)	2,412,502,632
Number of scaffolds	1,681
N_count	0.05%
N50	139,717,271
N50n	8
BUSCO (Genome)	97.80%
Gene models	74,088
Protein-coding genes	76,992
BUSCO (Proteome)	85.90%

C)



E)



ber	scaffold_length	scaffold_id	% of whole genome
	187,610,956	scaffold_1_1	7.78%
	178,682,930	scaffold_2_1	7.41%
	166,889,417	scaffold_3_1	6.92%
	162,452,925	scaffold_4_1	6.73%
	161,471,252	scaffold_5_1	6.69%
	157,427,254	scaffold_6_1	6.53%
	152,269,496	scaffold_7_1	6.31%
	139,717,271	scaffold_8_1	5.79%
	138,255,011	scaffold_9_1	5.73%
	127,884,699	scaffold_10_1	5.30%
	124,989,205	scaffold_11_1	5.18%
	120,464,456	scaffold_12_1	4.99%
	109,020,227	scaffold_13_1	4.52%
	108,330,513	scaffold_14_1	4.49%
	103,949,816	scaffold_15_1	4.31%
	2,139,415,428		88.68%
)			96.0% [S:77.4%,D:18.6%]

