High-quality genome assembly of *Cinnamomum burmami* (chvar. Borneol) provides insights into the natural borneol biosynthesis

Fangping Li[1, 2, 3, 5], Shilin Huang[3], Yu Mei[1, 6], Bingqi Wu[3], Zhuangwei Hou[4], Penglin Zhan[3], Shike Cai[1, 6], Qingmei Liu[5], Zhihao Hou[3], Junliang Zhao[2]*, Jihua Wang[1, 6]*

[1]Guangdong Provincial Key Laboratory of Crops Genetics & Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China.

[2]Rice Research Institute & Guangdong Key Laboratory of New Technology in Rice Breeding, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China.

[3]Guangdong Provincial Key Laboratory of Plant Molecular Breeding, State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou 510642, China.

[4]State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, 135, Xingangxi Road, Guangzhou, Guangdong, 510275, China

[5]Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou 510642, China;

[6]Guangdong Provincial Engineering & Technology Research Center for Conservation and Utilization of the Genuine Southern Medicinal Resources, Guangzhou 510640, China.

Correspondence: Jihua Wang[1, 6] (wangjihua@gdaas.cn), Junliang Zhao[2] (zhao_junliang@gdaas.cn)

Abstract: *Cinnamomum burmami* (chvar. Borneol) is a well-known medicinal and industrial plant cultivated in the Lingnan region of China. It is the key source from organism of natural borneol (D-borneol), one of the precious and widely used Chinese herbal medicines with a variety of medicinal effects. Here, we report a high-quality chromosome-scale genome assembly of *C. burmami* (chvar. Borneol) using Pacbio single-molecule sequencing and Hi-C technology. The assembled genome size was 1.14 GB with a scaffold N50 of 94.30 Mb, while 98.77% of the assembled sequences were anchored on 12 pseudochromosomes including 41549 protein-coding genes. Genomic evolution analysis revealed *C. burmami* and *C. micranthum* shared two Lauraceae unique ancestral whole-genome duplication (WGD) events. Likewise, comparative genomic analysis showed strong collinearity between these two species. Besides, the analysis for Long Terminal Repeat Retrotransposons (LTR-RTs) indicated the outbreak of LTR-RTs insertion made a great contribution to the size difference of genomes between *C. burmami* and *C. micranthum.* Furthermore, the candidate genes in pathway associated with natural

borneol synthesis were identified on the genome and their differential expressions were analyzed in various biological tissues. We considered that several of genes in Mevalonate (MVA) Methylerythritol Phosphate (MEP) pathways or in downstream pathway have the potential to be the key factors in the biosynthesis of D-borneol. We also constructed the genome database (CAMD; http://www.cinnamomumdatabase.com/) of *Cinnamomum* species for a better data utilization in the future. All these results will enrich the genomic data of Lauraceae plants and facilitate genetic improvement of this commercially important plant.

Key words: *C. burmami* (chvar. Borneol), Genome, D-borneol

Introduction

Natural borneol, also known as D-borneol, borneol, plum borneol, plum flower, etc., is a highly desirable natural product, as well as a precious Chinese herbal medicine. It has been widely used as folk medicine and food additives in China and India (Chen et al., 2010). Natural borneol has the effect of opening the orifice and awakening the mind, clearing away heat and relieving pain. It is often utilized for the treatment of multiple diseases such as sore throat, convulsive phlegm and shiver. Previous researches indicated D-borneol has remarkable effects on the aspects of anti-myocardial ischemia, brain tissue protection, nerve protection and anti-infection *in vivo* or *in vitro* (Armaka et al., 1999; Lu et al., 2010). Natural borneol is also a fragrance ingredient used in decorative

cosmetics products. Its use worldwide is in the region of 1–10 metric tones per annum (Bhatia et al., 2008). As of 2016, the annual demand for natural borneol in the world was about 12,000 tons, and the total annual output of natural borneol in China was less than 200 tons, indicating a huge market gap and potential (http://www.ibaogao.com/baogao/10241a2312016.html). As an important source of natural borneol, *Cinnamomum burmami* (chvar. Borneol) also called is a well-known medicinal and edible plant which had been widely cultivated in the Lingnan region of southern China (Wolf et al., 2011). Phytochemical studies have shown that *C. burmami* (chvar. Borneol) contains abundant natural borneol in its leaves (Huang et al., 2019). In China, Guangdong Province is the main planting area for *C. burmami* (chvar. Borneol), accounting for approximately 90% of its output (Chen et al., 2010).

Based on the identified metabolites and enzymes, previous studies have determined that the biosynthesis pathway of Glycerol 3-Phosphatase (GPP), the precusor of D-borneol *in vivo,* is related to the MVA pathway. Recently, MEP pathway in plants also had been reported to be associated with the GPP formation. Previous transcriptomic and synthetic biology studies also had identified that a gene (CbTPS) in TPS gene family is a key factor for downstream transformation of GPP generated by upstream pathways and natural D-borneol in *C. burmami* (Ma et al., 2021). However, due to the lack of a complete genome, systematic researches of D-borneol

89 biosynthesis and evolutiology of *C. burmami* at the genomic level were
90 still absent.
91 In order to fill this gap, we *de novo* assembled a chromosome-scale genome
92 of Mei Pian tree, a physiological type of *C. burmami* (chvar. Borneol) and
93 whose leaves is rich in D-borneol. The assembly was based on single-
94 molecule real-time long reads from the Pacific Biosciences (PacBio)
95 SequeII platform, and construct pseudo-chromosomes using high-
96 throughput chromosome conformation capture (Hi-C) techniques base on
97 a chromosome number of 2n = 24 as previous report (Wu et al., 1992).
98 By evolution and comparative genomic analysis, we detected that the
99 recent insertion outbreak of LTR-RTs was an important cause of plant
100 differentiation in *Cinnamomum*. RNA-seq transcriptiomics and further
101 gene annotation analysis based on our assembled *C. burmami* genome
102 dissected the natural D-borneol biosynthesis pathway at the level of
103 genome. The present study also demonstrated, *CbERG20* and *CbIDI1* in
104 MVA pathway, *CbDXR*, *CbCMS* and *CbHDR* in MEP pathway as well as
105 *CbTPS-1* in the downstream pathway, tend to be the key or the limiting
106 factor in the natural D-borneol biosynthesis pathway in *Cinnamomum*. To
107 encourage the data utilization, we also built a user-friendly genomic
108 database of *Cinnamomum* (CAMD;
109 http://www.cinnamomumdatabase.com/) for the data access.
110 The present study provides the first reference genome and detail genomic

information for *Cinnamomum burmami* (chvar. Borneol), a plant rich in D-borneol which is highly desired natural product used in medicine and industry. We also dissected the D-borneol synthesis pathway and provide a novel insight into the molecular mechanisms of this pathway. Our results will be valuable resources for promoting genetic improvement and understanding the biosynthesis of active ingredients of *C. burmami* (chvar. Borneol).

Method and materials

Plant materials and sequencing

Root, leaves, and stem samples of "Mei 3", a popular cultivated variety of *Cinnamomum burmami* (chvar. Borneol) were collected from the Mei Zhou City (Pingyuan District, 24°28′–24°66′ N, 117.3°13′–118.1°21′ E) in Guangdong Province, China. Because of its unique-geography and environment, the Pingyuan District of Mei Zhou City is considered the authentic production area of *C. burmami*. "Mei 3", with many excellent cultivation characteristics, including growth fast, high quality of component, and disease resistance, is the main variety of *C. burmami* (chvar. Borneol). In Pingyuan District. High-quality genomic DNA was extracted using a QIAGEN® Genomic Kit (QIAGEN, Germany). The integrity of the DNA was checked by 0.75% agarose gel electrophoresis.

Genomic survey and assembly

The genommic size and heterozygosis were estimated by the software GCE,

133 Jellyfish, kmergenie and Genomescope (Vurture et al., 2017). The data
134 generated form Pacbio SeqII was implemented to *De novo* assemble with
135 the software MECAT2 (Xiao et al., 2017) and polished by the software
136 NextPolish (Hu et al., 2019) combining the illumina short read data
137 mentioned in survey.

138 ***Hi-C experiment and chromosome assembly construction***

139 To further anchor the contigs to chromosomes, fresh young leaves was
140 utilized to construct a Hi-C library using a NEBNext Ultra II DNA Library
141 Prep Kit with the enzyme HandIII. 78.71 GB of clean data was generated
142 for Hi-C analyses by the Illumina NovaSeq 6000 platform. For the
143 anchoring of contigs, the mounting process was conducted by the pipeline
144 Juicer (Durand et al., 2016) and 3D-DNA (Dudchenko et al., 2017).
145 Software Juicerbox (Durand et al., 2016) was implemented to visualization
146 and adjustment.

147 ***Genomic assembly quality assessment***

148 Multiple methods were implemented to assess the accuracy and
149 completeness of the assembled genome. First, the paired-end reads were
150 mapped to the genome to evaluate its completeness using BWA-MEM with
151 the default parameters (Li, 2013). RNA-seq data from different tissues
152 (leaves, stalk, and root) were also aligned to the reference genome to obtain
153 the mapping rate using HISAT2 with the default settings (Kim et al., 2015).
154 Second, GC depth scatter plots were used to evaluate any contamination in

155 the sequencing data. Finally, the accuracy and completeness of the genome
156 assembly were evaluated by using BUSCOs to identify the single-copy
157 genes in the assembled genome with the Embryophyta_odb10 database
158 (Simão et al., 2015).

159 ***Gene prediction, Repeat element identification***

160 The sequence of repeat element in *C. burmami* genome was *de novo*
161 identified by software Repeatmodeler (Flynn et al., 2020) and masked by
162 Repeatmasker (Chen, 2004). The detection and taxonomy about the LTR-
163 RTs was accomplished by LTR_Finder (Xu and Wang, 2007), LTR_harvest
164 (Ellinghaus et al., 2008) and LTR_retriever (Ou and Jiang, 2018).

165 To generate gene annotation of *C. burmami* genome, software Augustus
166 was utilized in *de novo* gene prediction. A total of 65.74 GB of
167 transcriptome data from roots, stems, and leaves were alignment to the *C.*
168 *burmami* genome by Hisat2 and reassembled into transcripts by Stringties
169 (Pertea et al., 2015) and Transdecoder (Fangping et al., 2021; Kim et al.,
170 2015), which generated the site information of transcriptome. The protein
171 sequences of *C. micranthum* previously published (Chaw et al., 2019) were
172 utilized for the homologous gene prediction by GeMoMa (Keilwagen et al.,
173 2019). The information of *de novo* gene prediction, transcriptome and
174 homologous gene prediction was combining by the software EVM (Haas
175 et al., 2008). PASA and the transcripts generated by Transdecoder in
176 previous step were utilized to update the gff3 file for three rounds to add

177 alternatively spliced isoforms, adjusting the gene structure and add UTR
178 regions to gene models (Haas et al., 2008).

179 ***The Annotation of Gene function***

180 Functional annotation was achieved by using NCBI BLAST + v 2.2.2854
181 (Altschul et al., 1990) with cutoff e-values of 1e-5 and max target
182 sequences 20 to compare predicted proteins against public databases,
183 including SwissProt, Nr (Plant). The Best-hit BLAST results were then
184 considered as the gene functions. Gene Ontology (GO) identifiers for each
185 gene were obtained using InterProScan-5.25-64.0 by identify motifs and
186 domains by matching against public databases (Zdobnov and Apweiler,
187 2001). The predicted proteins were also submitted to KAAS to get KO
188 numbers for KEGG pathway annotation (M and S, 2000).

189 ***Phylogenetic analysis and Estimation of divergence time***

190 MAFFT v.7.271 (Katoh and Standley, 2013) with option --maxiterate 1000
191 was used to align 10 sets of Coding DNA Sequence (CDS) of 230 single-
192 copy orthologous groups. Each orthologous group alignment was used to
193 compute a maximum likelihood phylogeny using iqtree with the nucleic
194 acid substitution model GTR+F+R4 and 1000 bootstrap replicates
195 (Nguyen et al., 2015). The divergence time of each tree node was inferred
196 using MCMCtree of the PAML (Yang, 2007). The final species tree with
197 *Amborella trichopoda* as out-group and the concatenated translated
198 nucleotide alignments of 230 single-copy orthologues were used as input

199 of MCMCtree. The phylogeny was calibrated using various fossil records
200 or molecular divergence estimated by placing soft bounds at split node of
201 (MYA): *A. trichopoda* and *O. sativa* ('>1.73<1.99'), *D. carota* and *Salvia*
202 *splendens* ('>0.95<1.06'), *Arabidopsis thaliana* and *Populus trichocarpa*
203 ('>0.98<1.17') obtained from the TimeTree database (Kumar et al., 2017).

204 ***Synteny, Positive selection genes and WGD anaylsis***

205 The genomic alignment between *C. burmami* and *C. micranthum* was
206 finished and visualization by MUMMER (Kurtz et al., 2004). We
207 performed multiple sequence alignment to extract the conserved paralogs
208 of the protein sequences of this data by using BLASTP (E-value ≤ 1e−5).
209 MCScanX was used to identify collinearity blocks, and Ks values were
210 calculated by software WGDI to identify and extract collinear interval and
211 corresponding gene pairs (Sun et al., 2021). The calculation of Ks with the
212 method YN00 was implemented to predict the Whole Genome Duplication
213 (WGD) events

214 ***Transcriptome profiling analysis***

215 Pipeline Hisat2 was reads alignment of the reads from RNA-seq and the
216 read counts matrix was generated by software featureCounts (Liao et al.,
217 2014) for the quantitative analysis expression. R package DESeq2 was
218 implemented to normalize the read counts data and identified the DEGs
219 between different organs with Pvalue ≤ 0.01 and an absolute log2 (fold
220 change) ≥1 as the threshold (Love et al., 2014). Software STEM was

221 utilized to the cluster analysis DEGs expression and GO enrichment
222 analysis (Ernst and Bar-Joseph, 2006).

223 ***Tps gene family anaylsis***

224 Proteome data sets including *C. burmami* and other four sets *C.*
225 *micranthum, M. acuminate, O. sativa* and *D. carota* from NCBI were
226 utilized in TPS gene family analysis. Pfam domains: PF03936 and
227 PF01397, were used to identify against the proteomes using HMMER
228 (version 3.0; cut-off at e-10) (Finn et al., 2010; Finn et al., 2011). Putative
229 or annotated protein sequences of *TPS* were aligned using MAFFT with
230 default parameters and manually adjusted using MEGA. The TPS gene tree
231 was constructed using iqtree with 1,000 bootstrap replicates. The
232 subfamily *TPS*-a was designated as the out-group. Branching nodes with
233 bootstrap values of <80% were treated as collapsed.

234 ***Prediction of protein structure modeling and molecular docking***

235 Homology modeling was performed with Rosettafold with default *de novo*
236 prediction parameter (Baek et al., 2021). The results were inspected and
237 rendered with PyMOL (DeLano, 2002). Protein docking and binding
238 energy calculation were done with AutoDock Vina using local search
239 parameters and default docking parameters (Trott and Olson, 2010). The
240 alignment and the visualization of amino acid sequences was finished by
241 platform ClustalW and ESPript (Gouet et al., 1999).

242 ***Data and code availability***

243 The sequencing data, including the assembly data of *C. burmami* was
244 submitted to the National Genomics Data Center (NGDC) under
245 BioProject accession number PRJCA006286. All the assembly and
246 annotation data also can be accessed from CAMD
247 (http://www.cinnamomumdatabase.com/)
248 Customized codes used in this study have been deposited in GitHub
249 https://github.com/lifangpings/CB_genome.

250 ***Real-time Quantitative PCR (q-PCR) comfirmation***

251 For q-PCR assays, the total RNA was extract using Aurum Total RNA Mini
252 Kit (BIO-RAD) and cDNA was obtained by thermo scientific RevertAid
253 First Strand cDNA Synthesis Kit according to the supplier's protocols
254 respectively. All the primers' sequence was obtained from the National
255 Centre for Biotechnology Information (NCBI) database. According to kit
256 instruction 10 μl reaction system was prepared and q-PCR was operated
257 using CFX96 Real-Time System (BIO-RAD). For data analysis, all
258 detected genes expression were normalized against Actin-F
259 (CBB03T203270) of *C. burmami*.

260 ***Database construction.***

261 All genomic sequence, annotation, were stored via MySQL on a Centos7
262 server. A user-friendly website was developed using HTML5, JavaScript
263 and PHP7, which can be accessed through different browsers, such as
264 Google Chrome and Firefox. Gene models and transcript isoforms were

265 provided via JBrowse (Skinner et al., 2009; Buels, R. et al., 2016). The
266 query searches were achieved via PHP7. Common utilities for genomic
267 studies such as BLAST, and Sequence exactor were also deployed and
268 accessible.

269 **Result**

270 ***Genomic assembly***

271 To evaluate the genome size and heterozygosity of *Cinnamomum burmami*
272 (chvar. Borneol), a total of 61.4 GB short reads from the MGISEQ-2000
273 sequencing platform were subjected to K-mer analysis. The 17-mer
274 frequency curve showed a unimodal distribution, with the highest peak
275 occurring at a depth of 54 (Supplementary Fig. S1a). Based on the total
276 number of K-mers, the genome size and heterozygosity of *C. burmami*
277 (chvar. Borneol), were estimated to be 0.94-1.25 GB and ~0.67%,
278 respectively (Supplementary Fig. S1b; Supplementary Table. 1).

279 A total of 170.1 GB single molecule sequencing long read data with the
280 average length about 32 kb (N50=21.38kb) was utilized in the genomic *de*
281 *novo* assembly. After polishing combining the short-read data and
282 removing redundant/contaminated sequences, the size of the final contig
283 was 1.15 GB with a contig N50 of 1.13Mb (Table.1). A Hi-C interaction
284 heatmap indicated 12 highly obvious interacted range of the signal which
285 is corresponding with the previous research about the chromosome number
286 of *C. burmami* (chvar. Borneol) (Supplementary Fig. S2). Based on the

287 signal, a total of 98.77% of the assembly was anchored to 12 pseudo-
288 chromosomes. The contig sequences were connected in the determined
289 order and direction by adding 100 N to obtain the final chromosome-level
290 genome sequence with a chromosome mount rate of 98.77% the size
291 ranged from 66.1Mb to 140.5Mb, and with Scaffold N50 = 94.90Mb. The
292 ordering, and orientation of the contigs was also arranged, providing the
293 first high-quality chromosome-scale genome assembly for *C. burmami*
294 (Table.1).
295 Multiple method was utilized for the quality assessment of the novel *C.*
296 *burmami* genome. The method of Benchmarking Universal Single-Copy
297 Orthologs (BUSCOs) was implemented to evaluate the assembly quality
298 of *C. burmami* genomes, which identified 89.7% of the complete BUSCOs
299 in the assembly by using the reference of the conservative database of
300 dicotyledons (Supplementary Table S2). For the assessment of genomic
301 consistency, the short reads were mapped to the genome using bwa
302 software. This analysis indicated ~99.78% of the clean data were mapped
303 to the genome assembly. The mapping ratio of the RNA-seq reads from
304 different tissues was in the range of ~92.08%–94.93%. The GC content of
305 this genome is 39.02% while the sequencing depth was concentrated at
306 110–150×. All these methods of assessment elucidated a high completeness
307 and consistency of the *C. burmami* genomic assembly (Table.1).
308 ***Repeat element identification and Genome annotation***

309 The *de novo* prediction based on RepeatModeler indicated 673.32 Mb of
310 repetitive sequences, accounting for 63.24% of the assembled genome.
311 Combining the result of *de novo*, homolog-based, and transcriptome-based
312 predictions, a total of 41549 non-redundant protein-coding genes were
313 generated across 12 chromosomes (Supplementary Fig. S3). The average
314 gene length and coding sequence size were 2962.35 and 1142.14 bp,
315 respectively, with an average of five exons per gene. The result of function
316 annotation indicated that 37799 genes (90.97 %) was annotated in at least
317 one pathway of the public database. Further functional annotation using
318 InterProScan estimated that of the genes 33582 (80.82%) contained
319 conserved protein domains recorded by Panther and Pfam database. 54.38%
320 of the genes could be classified with Gene Ontology (GO) terms. 8401
321 (20.22%) of the genes mapped to known plant biological pathways based
322 on the KEGG Pathway database.

323 ***Phylogenetic analyses***

324 To identify the phylogenetic relationship and position of *C. burmami* and
325 *C. micranthum* we constructed a phylogenetic tree based on 230 strictly
326 single-copy orthologous sets. Consistent with expectations, phylogenetic
327 tree indicated clustering of two species of *Cinnamomum* on one branch.
328 The maximum likelihood (with a high bootstrap index) and MCMCtree
329 with fossil calibrations analysis indicated that the divergence between
330 magnoliids and eudicots tended to happen in 201 Mya (95% confidence

interval: 196.1-209.1). The time of divergence between *C. micranthum* and *C. burmami* to be 8 Mya (95% confidence interval: 7.5-8.5) (Figure .1).

LTR-RTs promote species differentiation among Cinnamomum

In the identification of LTR-RTs for these two species, a total of 186.62Mb (25.1% of the whole genome) LTR-RTs sequence was detected in *C. micranthum* while 459.42Mb (40.59% of the whole genome) was detected in *C. burmami* (Figure. 2a and Supplementary Table S2). The insert time analysis elucidated these two species experienced a significant recent outbreak of LTR-RTs insertions. However, much more insertions of LTR-RTs were speculated to occur in 0-8Mya in *C. burmami* than in *C. micranthum,* which was also corresponding with the divergent time between these two species. These LTR-RTs were also evenly distributed in the genome of *C. burmami* (Figure. 2b).

This ~250 Mb length difference of LTR-RTs tends to be a preliminary explanation for the interspecific genome size differences in *C. micranthum* and *C. burmami* genera.

Whole-genome duplication and Synteny analysis

To more precisely infer the timing of the two rounds of WGD evident in the *Cinnamomum,* the intragenomic and interspecies homologue *Ks* (synonymous substitutions per synonymous site) distributions were estimated with the genome of *C. micranthum, C. burmami* and *A. coerulea* (as outgroup). The curve of *Ks* distribution indicated that two peaks at ~0.5

353 to 1.5 in both *C. micranthum* and *C. burmami* genomes, which elucidated
354 these two species shared the same two ancient WGD events (Two ancestral
355 WGD events endemic to Lauraceae before divergence of eudicots and
356 magnoliids) before their divergence and no recent independent WGD event
357 was detected (Figure. 2c).
358 Consistent with the predictability, both the results of whole genome
359 comparison and gene collinear comparison between *C. micranthum* and *C.*
360 *burmami* indicated an obvious synteny presence (Figure. 3a and
361 Supplementary Fig. S4). Based on the collinearity of *C. burmami* and *C.*
362 *micranthum* genomes, we sorted and renamed *C. burmami* genome
363 sequences. We found that the relative proportions of chromosome sizes
364 were consistent between the two species (Table. 2). Besides, large
365 fragments of chromosomal translocations were displayed in Chr2, Chr 4,
366 Chr5, and Chr12, while a large fragment inversion occurs in front of Chr6
367 in comparison between this two *Cinnamomum* species. This phenomenon
368 suggested that the translocations and inversion of chromosomes might
369 have been occurred during the process of *Cinnamomum* species
370 differentiation.
371 ***Characteristic analysis of genes showing organ-specific expression***
372 A total of 31537 genes were detected to expression in the organs of *C.*
373 *burmami*. The gene expression levels analysis indicated that 1318, 1208,
374 672 genes expressed explicitly in roots, stalks, leaves respectively, while

375 24681 genes were expressed in all tissues (Figure. 3b).

376 To elucidate the similarities and differences of gene expression patterns in
377 different tissues, we performed a k-means cluster analysis. A total of 20444
378 differentially expressed genes (DEGs) were divided into 10 clusters. We
379 focused on clusters containing tissue-specific expressed genes, particularly
380 those that were stable in roots and stalks while increased in leaves (3348
381 genes in Cluster1) (Figure. 3c). Apart from the genes involving in
382 photosynthesis, genes in the terpene synthase (*TPS*) activity pathway (GO:
383 0010333) were shown to be expressed and enriched in leaves, which was
384 consistent with the higher content of terpenoids in leaves including D-
385 borneol in *C. burmami* reported by previous studies. Besides, a number of
386 methylation related genes (GO: 0004222 and GO: 0008173) were also
387 significantly increased and enriched in leaves. This phenomenon suggested
388 the regulation associated with epigenetic may play a role in development
389 and compound synthesis in the leaves of *C. burmami* (Figure. 3d).

390 ***TPS gene family analysis***

391 Similar to another plant of *Cinnamomum, C. micranthum*, a large number
392 of *TPS* genes (*CbTPS*) located in the genome of *C. burmami.* A total of 73
393 *CbTPS* genes were predicted and annotated. Phylogenetic analysis of
394 *CbTPS* placed these genes into six *TPS* gene subfamilies that have been
395 described for seed plants. Among this *CbTPS* genes, 1 and 2 were placed
396 in the *TPS-c* and *TPS-e* subfamily respectively. These two subfamilies have

397 been reported to encode diterpene synthases like copalyl diphosphate
398 synthase and *ent*-kaurene synthase (Chaw et al., 2019). These are key
399 enzymes catalyzing the formation of the 20-carbon isoprenoids
400 (collectively termed diterpenoids; C20s). The remain *CbTPS* genes
401 attributed to *TPS*-a, *TPS*-b, *TPS*-f and *TPS*-g, which probably encode the
402 10-carbon monoterpene (C10) synthases, 15-carbon. With 29 and 23
403 homologues, respectively, *TPS*-a and *TPS*-b subfamilies are most diverse
404 in *C. burmami*, presumably contributing to the mass and mixed production
405 of volatile C15s and C10s, including D-borneol (Figure. 4 and
406 Supplementary Table S4).
407 Besides, *CbTPS* genes are not uniformly distributed throughout the
408 chromosomes and clustering of members from individual subfamilies was
409 observed as tandem duplicates. Among these genes, forty three copies
410 (58.9%) of *CbTPS* genes belonging to different subfamilies were found in
411 the 53–59Mb region of chromosome 6, and 56–61Mb region of
412 chromosome 9. Two to nine *TPS* family genes were detected in tandem on
413 chromosomes 1, 3, 4, 7, 10 and 11 while only one gene of *TPS* family was
414 detected in chromosomes 5. chromosomes 7 contains 29 *CkTPS* genes
415 belonging to several subfamilies, including all of the eight *CkTPS*-a, 12
416 *CkTPS*-b, five *CkTPS*-e and three *CkTPS*-f Genes belonging to this cluster
417 were not grouped together in their corresponding subfamily phylogen,
418 suggesting that their arrangement might have occurred more recently than

419 the last WGD event (Supplementary Fig. 4b).

420 ***Identification of genes related to the D-borneol biosynthesis pathway***

421 Previous research indicated that MVA pathway play a key role in the
422 upstream biosynthesis pathway of D-borneol. Our gene annotation results
423 indicated that all MVA pathway genes (*CbERG10, CbERG13, CbtHMG1,*
424 *CbERG12, CbERG8, CbERG19, CbIDI1, CbERG20*) distributed in 12
425 chromosome of *C. burmami.* The transcriptome analysis elucidated the
426 expression level of genes *CbERG20, CbIDI-1,* was significantly elevated
427 in leave comparing to stem and root. *CbERG20* is the key enzyme to
428 promote the synthesis of GPP, which is a directly related raw material of
429 D-borneol synthesis.

430 Besides, MEP pathway also have been reported to participate in the
431 precursor GPP formation in the D-borneol formation process. In this
432 pathway, the expression of multiple genes such as *CbCMK, CbMCS* and
433 *CbHDR* showed similar expression pattern with above key genes in MVA
434 pathway. (Figure. 5).

435 For the downstream pathway, *CbTPS-1* (*CBB06T352390*), encoded by a
436 gene of *TPS*-b gene subfamily, has been reported to directly mediate the
437 transformation of GPP to D-borneol (Ma et al., 2021). The expression
438 levels of these key genes for D-borneol synthesis were much higher in
439 leaves than in stem or root implied they may be the key or the limiting
440 factor in the biosynthesis of D-borneol in *C. Burmami* leaves (Figure. 5).

441 Gene cloning and qPCR analysis to above potential key genes were further
442 conducted, and results are highly consistent with our transcriptomic and
443 gene annotation results, indicating the accuracy and reliability of our
444 results (Supplementary Fig. S6).

445 ***Sequences differences cause configuration alteration of the docking site***
446 ***of CbTPS-1 and CkTPS-1***

447 The 3D structure model of protein generated by Rosettafold was
448 implemented to model the protein structure of *CbTPS-1* and its highly
449 homologous *CkTPS-1*(*RWR98053.1*) in *C. micranthum*, and molecular
450 docking analysis with small micromolecule GPP was also performed. This
451 analysis showed an obvious GPP binding pocket in the middle of these two
452 *TPS* proteins folding structure (Figure. 6a). The GPP docking pocket
453 located in the ~300aa to ~500aa of these two TPS protein. A total of 26
454 amino acid mutations were detected in this region. Among these, 15
455 mutations changed the polarity of the amino acids at that location, which
456 have potential to affect their protein folding (Figure. 6b). Therefore, due to
457 folding differences caused by changes in amino acid sequence, GPP can
458 form more hydrogen bonds with more A-helices (4 > 1) in *CbTPS-1*
459 binding pockets (9 > 8) and its binding energy in *CbTPS-1* is also higher
460 than that of *CkTPS-1* (-7.7<-7.4). The higher binding capability of *CbTPS-*
461 *1* may cause of higher D-borneol production in *C. Burmami* than other
462 These differences may be the main cause of *C. Burmami* to produce higher

463 D-borneol than other *Cinnamomum.* plants (Figure. 6c).

464 ***Construction of genome database of Cinnamomum species***

465 To meet demand for pitaya genome and multiomics data resources, we
466 established an integrated pitaya genome and multiomics database (CAMD;
467 http://www.CinnamomumDatabase.com). The CAMD provides a
468 comprehensive consultation service including genomic blast, gene browser
469 and sequence fetch with the data from the latest assemblies of *C. burmami*
470 and *C. micranthum* genomes (Chaw et al., 2019) (Figure. 7).

471 **Discussion**

472 *Cinnamomum*, a genus of about 250 species of the Lauraceae, is widely
473 distributed in tropical Asia, Australia to Pacific islands, and tropical
474 America. The plants of *Cinnamomum* are rich in volatile oil, terpenoids,
475 tannins and their analogues, aromatic compounds and polysaccharides.
476 Multiple species of this genus are important raw materials in chemical
477 industry and medicine. However, the only one assembled genome in this
478 species is *C. micranthum* (Chaw et al., 2019) . There are few reports on
479 further genome research of *Cinnamomum* genus. The lack of high-quality
480 genome and detailed genomic information has limited our understanding
481 of molecular mechanisms underlying valuable natural product and genetic
482 improvement of these plants.

483 In the present study, we assembled a chromosomal scale, high-quality
484 genome of *Cinnamomum burmami* (chvar. Borneol), as well as detailed

genomic and transcriptomic information. Compared with the first published genome of *Cinnamomum species, C. micranthum,* the assembly of *C. burmami* achieved much higher quality, such as a longer sequence length of contigs (1.14 Mb > 498.92Kb) and a higher chromosome mount rate (98.31% > 90.1%) with a larger genomic size (1.13Gb > 730.7 Mb). These improvements suggested that the assembly of *C. burmami* tends to be a valuable and better reference genome for future genomic study of *Cinnamomum* species.

The construction of the *C. burmami* genome and the comparative genome analysis also shed light on the understanding of genome evolution of Lauraceae. Our results confirmed that Lauraceae and camphor plants shared two ancestral WGD events. As the representative of magnoliid, the phylogenetic relationship analyses of *C. burmami* also resolve magnoliids to be closer to eudicots than to monocots. Furthermore, the outbreaks of LTR-RTs have been considered to be an important impetus of increasing genetic diversity of multiple species (Sun et al., 2020; Lin et al., 2018). The combination of LTR-RTs size and insertion time indicated that recent LTR-RTs insertion events also promoted the differentiation between species *C. micranthum* and *C. burmami* in *Cinnamomum.*

Our assembled genome and transcriptomic analysis results also help to elucidate the molecular mechanism of borneol biosynthesis. As a medicinal plant, *C. burmami* (Chvar. borneol) is an important source of D-borneol.

507 D-borneol, which is also called natural borneol, is one of the most famous
508 Chinese traditional medicines. It has the effect of awakening, clearing heat
509 and relieving pain. It is often utilized popularly for fever, dizziness,
510 convulsion, stroke, phlegm, sore throat, purulent ear canal and other similar
511 symptoms. Since its discovery in the 1980s, *C. Burmami* (Chvar. borneol),
512 which is rich in natural borneol, has been widely grown as a medicinal crop
513 in city Meizhou in Guangdong province of China. It is very important to
514 fully elucidate the molecular pathway of D-borneol biosynthesis for further
515 natural plant utilization and future synthetic biology application. Ma et.al
516 had reported the reconstitution of D-borneol biosynthetic pathway in
517 *Saccharomyces cerevisiae*, which is a critical step for borneol biosynthesis
518 application. However, genes responsible of borneol synthesis, and theirs
519 underlying regulation mechanisms in plants are still largely unknown.
520 With the help of high-quality reference genome of *C. burmami* (Chvar.
521 borneol) assembled in the present study, we were able to fully dissect the
522 natural borneol synthesis pathway from a genome scope. Previous studies
523 have confirmed MVA pathway is the key pathway for the synthesis of GPP,
524 which serve as the precursor of D-borneol (Ma et al., 2021). The MEP
525 pathway has also been reported to be involved in GPP synthesis in plants
526 (Banerjee and Sharkey, 2014). The transformation of GPP and D-borneol
527 was completed with the participation of *TPS* family, while *CbTPS-1* has
528 been confirmed to participate in D-borneol biosynthesis directly (Ma et al.,

529 2021). In the present study, all genes participate in the both natural borneol
530 biosynthesis pathway were identified in *C. burmami* (Chvar. borneol)
531 genome. Transcriptomic analysis showed key genes in both pathways tends
532 to highly expressed in leaves, rather than in stem or root. All these results
533 provided important genomic information of genes for the following
534 research about D-borneol biosynthesis in *C. burmami*.
535 Furthermore, the analysis of *TPS* gene family elucidated multiple *TPS* gene
536 clusters were detected in *C. burmami* genome, especially in Chr6 and Chr9.
537 Recent studies have shown that several plants biosynthetic gene clusters
538 (BGCs) of natural products exist in plant genomes. These plant
539 biosynthetic gene clusters generally participate in the process of natural
540 product synthesis synergistically (Liu et al., 2020; Mao et al., 2020). In the
541 present study, the function and mechanisms of BGCs formed by *TPS*
542 family genes are worth more profound exploration.
543 Additionally, our transcriptomic analysis further identified genes
544 participate in methylation process were significantly enriched in genes
545 specific expressed in leaves, but not in stem or root. Several researches
546 have reported the epigenetic processes, such as DNA or RNA methylation
547 played important roles in the synthesis of secondary metabolites in
548 multiples species (Zuo et al., 2020; Wang et al., 2019). Leaves is the major
549 organ for borneol synthesis in *C. burmami*. Together with the results that
550 expressions of genes of borneol biosynthesis pathway were much higher in

leaves than in stem and root, and the results that methylation related genes enrichment in leaves specific expression genes, implied methylation may play important roles in the expression regulation of gene related to borneol synthesis in *C. burmami* leaves. Methylation and gene expression regulation may be further leverage as target for borneol production improvement in *C. burmami* (Chvar. borneol). The mechanisms behind these results need further investigation.

In the present study, we also constructed a genome database (CAMD), which combined our results of *C. burmami.* and genome information from another *Cinnamomum* plant, *C. micranthum.* (Chaw et al., 2019). CAMD is the first database of the *Cinnamomum* and will become an excellent central gateway to better understand the biology and genetics of Lauraceae. The release of CAMD will benefit for studies of genetic diversity and improvement in *Cinnamomum*.

Conclusion

Here we presented a high-quality complete genome, as well as detailed genomic information of *C. burmami* (Chvar. borneol). The high-quality genome provides novel insights about genome evolution of *C. burmami* , as well as *Cinnamomum* genus. With the help of assembled and annotated genome, we dissected the pathways of borneol synthesis, and identified key genes in these pathways. Transcriptomic analysis further shed light on the further investigation of regulation and molecular mechanisms of these

pathways. We also constructed the genome database (CAMD) of *Cinnamomum* species for better data utilization in the future. The availability of *C. burmami* complete genome and its genomic information from this work will provide a valuable genome resource for understanding the evolution, active ingredient biosynthesis, and genetic improvement for not only *C. burmami* itself, but also *Cinnamomum* genus.

Acknowledgements

This work was supported by the scientific innovation strategy-construction of high-level Academy of Agriculture Science (R2019PY-JX003 and R2019PY-JX001), Natural Science Foundation of Crops Research Institute, Guangdong Academy of Agricultural Sciences (0145). We thank Prof. Wang Shaokui (Department of Crop genetics and breeding, South China Agricultural University) and Prof. Zhou Xiaofan (Integrative Microbiology Research Centre, South China Agricultural University) for kindly help and support.

Author contributions

LF, ZJ and WJ conceived and designed this research. LF, HS, XS, WB, HZ, ZP, MY, ZJ, LQ, HZ and WJ performed the experiments. LF, HS, WB, ZJ and WJ prepared the article.

Competing interests

The authors declare no conflict of interest.

Figure

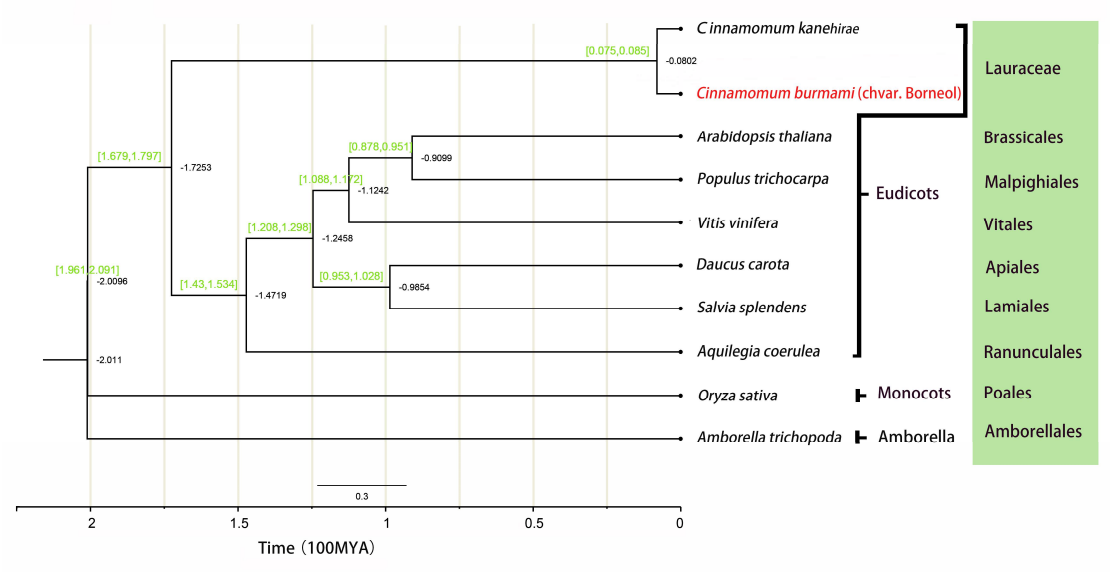

Figure.1

The phylogenetic relationship of *C. burmami, C. micranthum* and other eight plant species of different family. The green numbers in the brackets denote divergence time estimates. All nodes' bootstrap support was 100 unless stated otherwise.

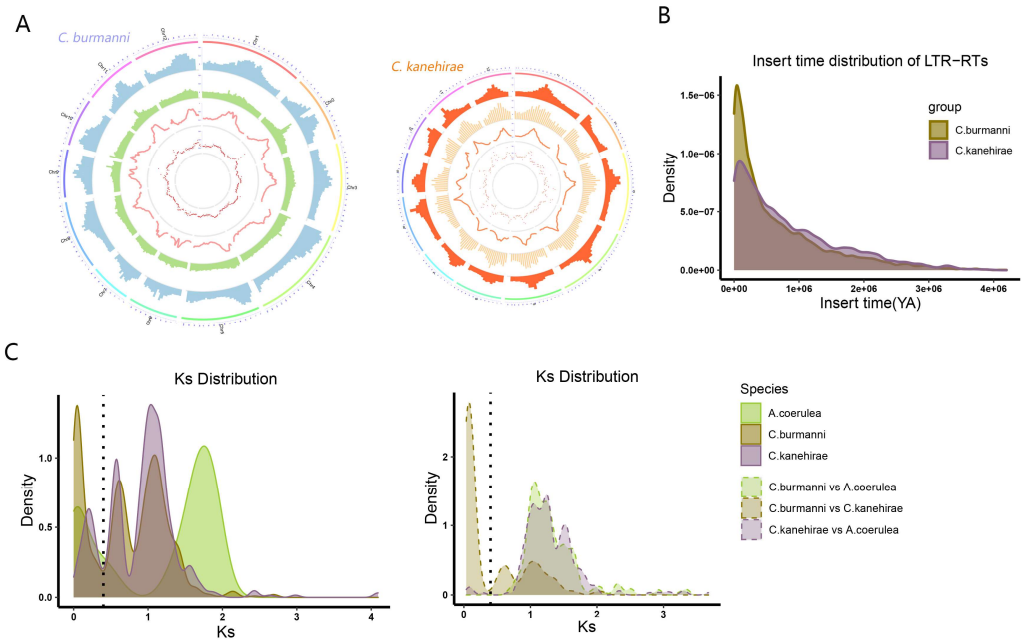

Figure. 2

The distribution and insertion times of the LTR-RTs in *C. burmami* and *C. micranthum* genomes as well as the WGD event identification. **A:** LTR-RTs Characterization of the *C. burmami* and *C. micranthum* genome; the circle from outside to inside represents chromosomes LTR-Copia density, LTR-Gypsy density, Total LTR density and GC content All distributions are drawn in a window size of 3 Mb. **B:** Quantity and time distribution of LTR-RTs **C:** Ks distribution among *C. burmami*, *C. micranthum* and *A. coerulea* Lines show Ks distribution within (continuous) and between genomes (dashed).

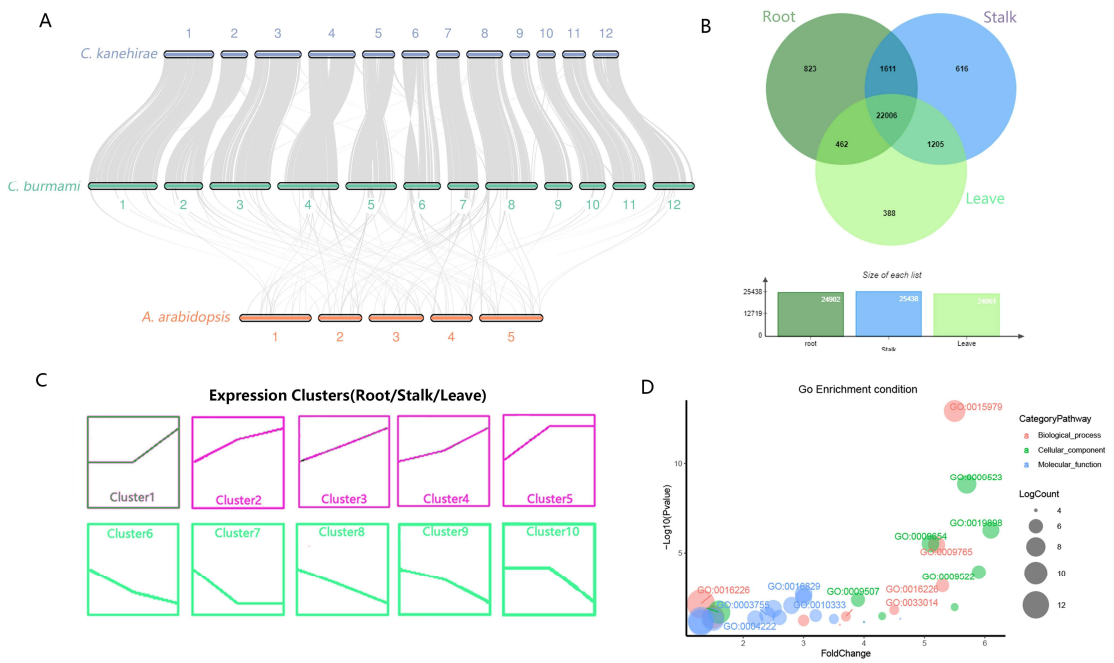

Figure. 3

Collinearity analysis between *C. burmami, C. micranthum* and *A. Arabidopsis* as well as the gene's organ-specific expression of different organs *of C. burmami.* **A:** Gene Collinear relationship of *C. burmami* and *C. micranthum* and *A. arabidopsis*. The gray line connects matched gene

pairs. **B:** The Venn diagram visualized the expressed genes in roots, stalks, and leaves. The overlapping regions represent genes expressed in at least two tissues, while the non-overlapping regions represent tissue-specific genes. **C:** Hierarchical clustering showing the expression patterns of DEGs in roots, stalks, and leaves.

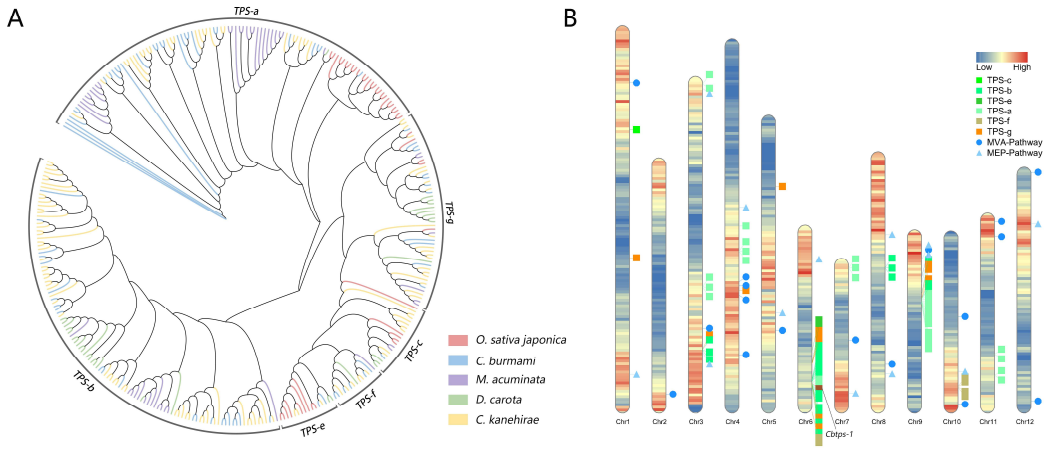

Figure. 4

The gene density and the gene distribution involving in *TPS* gene family and D-borneol formation; Phylogenetic placements of the 73 *TPS* family genes of *C. burmami*. **A:** The phylogenetic tree was constructed using putative or characterized *TPS* genes from 5 plant genomes including *C. burmami*. **B:** Gene density distribution as well as the location of the gene involving in *TPS* gene family and D-borneol formation.

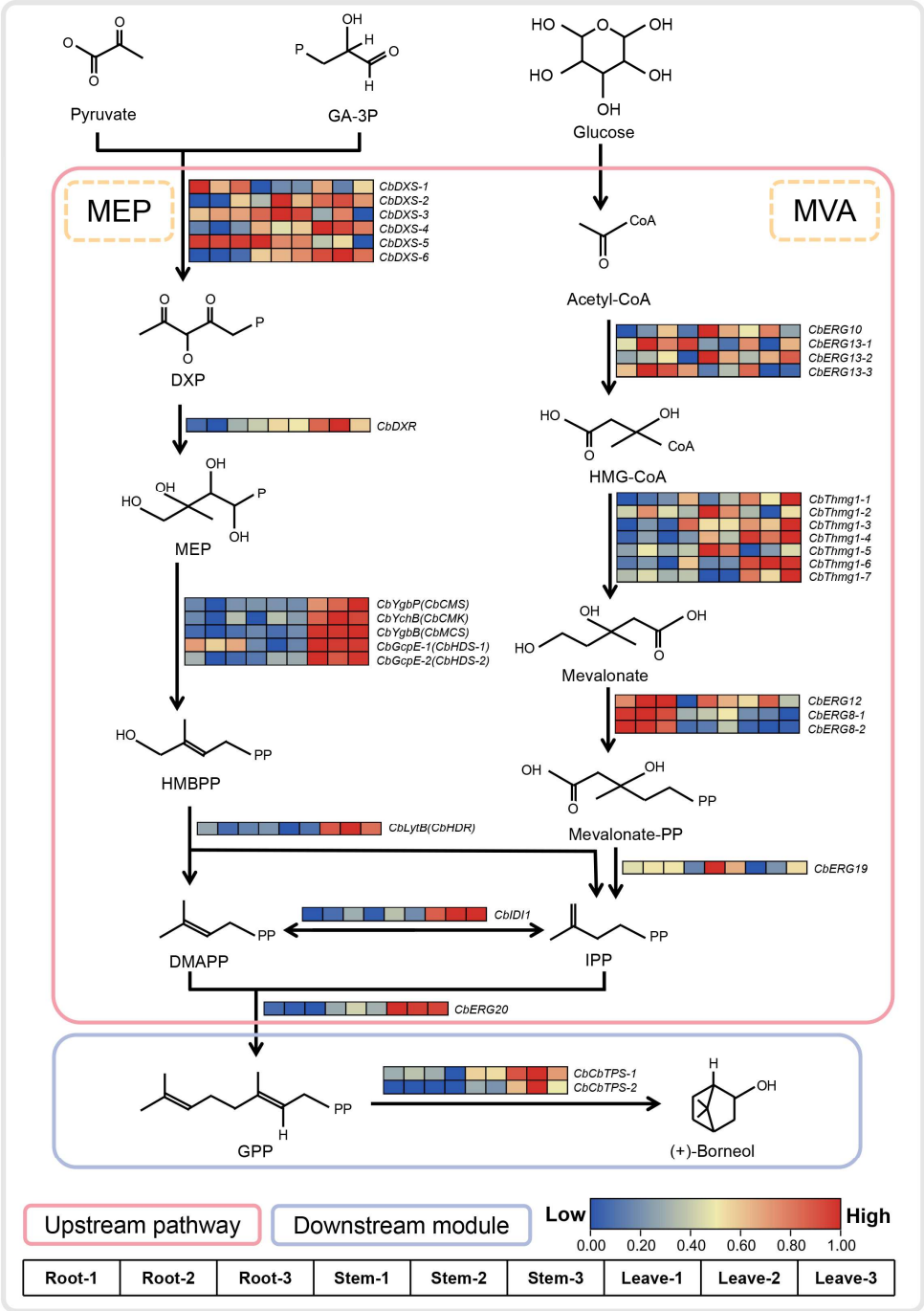

Figure. 5

Identification and expression analysis of genes involved in MVA and MEP pathway for key material GPP production (pink ellipses) nd the biosynthetic pathway of D-borneol (blue ellipses). Heat maps show the expression of genes involved in these pathways

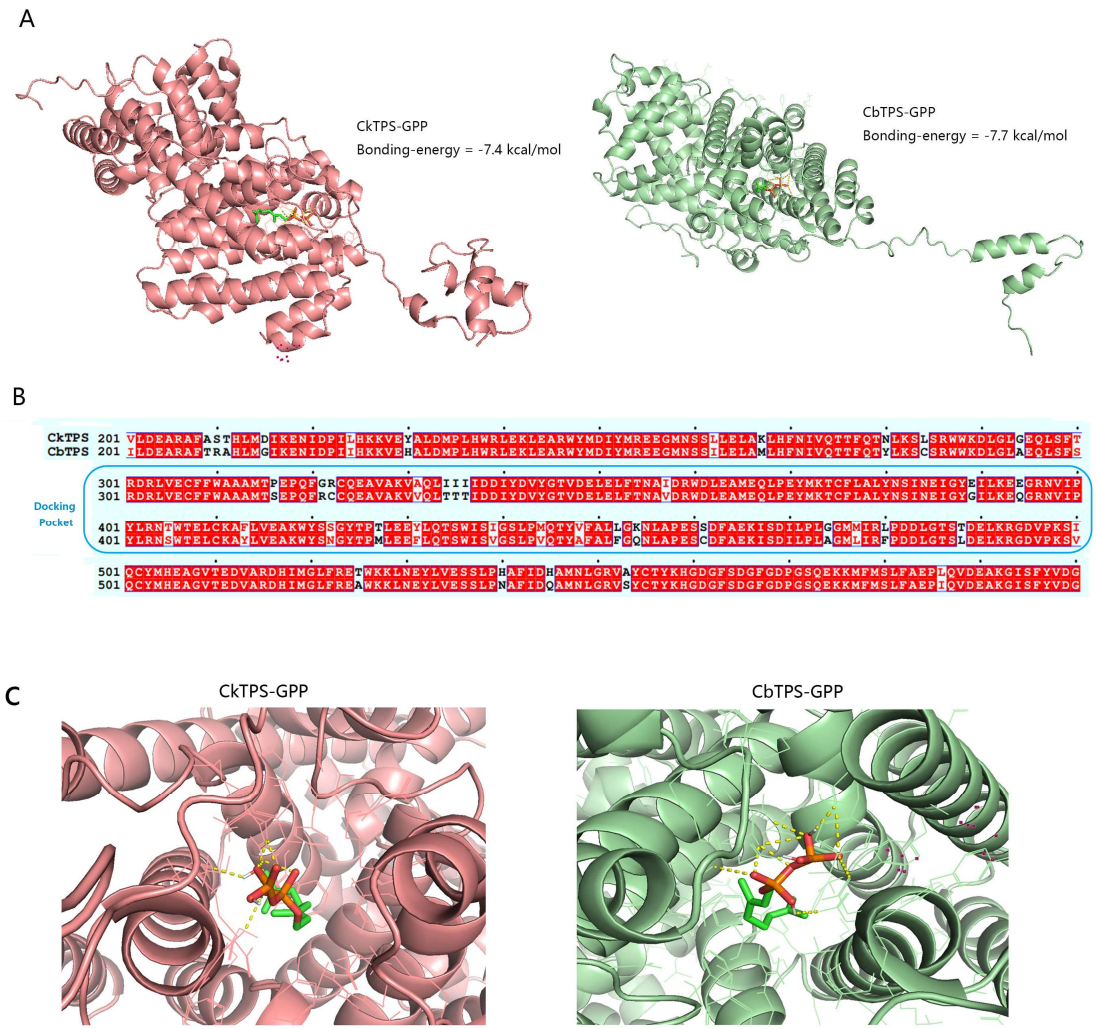

Figure. 6

Molecular docking result between *CbTPS-1* and its highly homologous gene *CkTPS-1* in *C. micranthum.* **A:** 3D structure of proteins and the overall docking situation (docking pocket position; *CkTPS-1*-GPP/*CbTPS-1*-GPP); **B:** The sequences comparison of Amino acid docking pocket position **C:** The details of docking (*CkTPS-1*-GPP/*CbTPS-1*-GPP).

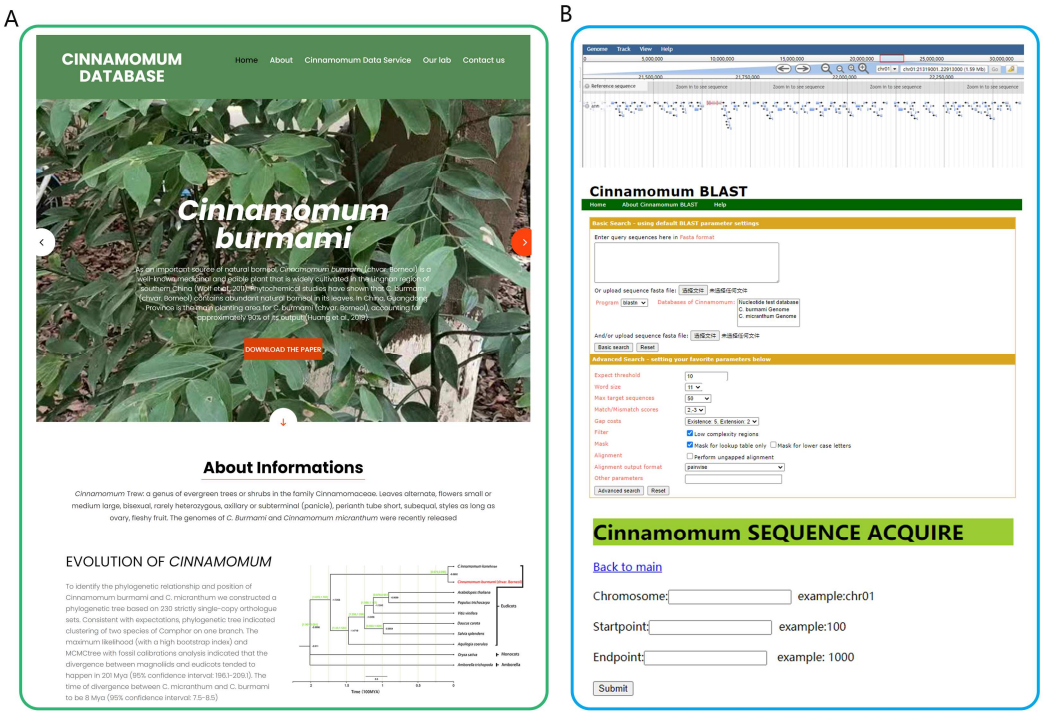

Figure. 7

The screenshots of representative resources and action page example of CAMD; **A:** Main pages of CAMD; **B:** Partial function pages displayed including Genomic blast, Gene Brower and sequence fetch.

652 **Table.1** Summary and comparison of *C. burmami* and *C. micranthum*

653 assembly

Statistics without reference	*C. burmami* (contig)	*C. burmami* (Chromosome)	*C. micranthum* (Chromosome)
Contigs	4653	76	2150
Contigs (>= 50000 bp)	2109	76	268
Largest contig	5837401	140537727	87013940
Total length	1154926090	1132685329	730416403
Total length (>= 0 bp)	1154926090	1132685329	730416403
Total length (>= 50000 bp)	1115042729	1132685329	701509862
N50	1144013	94903400	50350591
N75	484956	72773248	45778809
L50	292	5	6
GC (%)	39.15	39.02	38.34

654

655 **Table.2** Statistics of the sizes of each chromosome of *C. burmami* and *C.*

656 *micranthum*

Chromosome	*C. burmami*	*C. micranthum*	Corresponding Chromosome No. (*C. micranthum*)
Chr1	140537727	87013940	QPKB01000001.1
Chr2	92646709	50350591	QPKB01000006.1
Chr3	122226000	81065106	QPKB01000002.1
Chr4	135862046	77926444	QPKB01000003.1
Chr5	108037362	58596509	QPKB01000005.1
Chr6	68311217	48047431	QPKB01000007.1
Chr7	56091752	39555212	QPKB01000011.1
Chr8	94903400	58831143	QPKB01000004.1
Chr9	66575986	41984183	QPKB01000010.1
Chr10	66169134	37069385	QPKB01000012.1
Chr11	72773248	45778809	QPKB01000009.1
Chr12	89292919	46631104	QPKB01000008.1

657

658

List of Supplementary materials

Supplementary Figure

Supplementary Figure.1 Evaluation of genome size and heterozygosity

Supplementary Figure.2 HI-C heatmap of *C. burmami*

Supplementary Figure.4 Collinearity of genomic sequences between *C. burmami* and *C. micranthum*

Supplementary Figure.5 qPCR confirmation of partial gene involving in D-borneol biosynthesis in *C. burmami*

Supplementary Figure.6 Gene clone of *CbERG20* and *CbTPS-1* from *C. burmami*

Supplementary Table

Supplementary Table.1 Genomic survey result of *C. burmami* generated by software Gce

Supplementary Table.2 Busco assessment results of *C. micranthum*

Supplementary Table.3 Length and types Statistics of LTR-RTs

Supplementary Table.4 Numbers of *TPS* subfamilies in the 5 genomes

Supplementary Table.5 List about the gene in *TPS* family or involving in D-borneol biosynthesis in *C. burmami*

Supplementary Table.6 Sequence and List of primers in qPCR and gene clone confirmation assays.

682 **References:**

683 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman DJ (1990)
684 Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

685 Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee,
686 G.R., Wang, J., Cong, Q., Kinch, L.N. and Schaeffer, R.D. (2021)
687 Accurate prediction of protein structures and interactions using a
688 three-track neural network. *Science* **373**: 871-876.

689 Buels, R., Eric, Y., Colin, M.D., Richard, D.H., Monica, M., Gregg, H.,
690 David, M.G., Christine, G.E., Suzanna, E.L., Lincoln, S. and Ian, H.
691 H. (2016) JBrowse: a dynamic web platform for genome visualization
692 and analysis. *Genome Biol* **17**: 66.

693 Banerjee, A. and Sharkey, T.D. (2014) Methylerythritol 4-phosphate (MEP)
694 pathway metabolic regulation. *Nat Prod Rep* **31**: 1043-1055.

695 Bhatia, S.P., Letizia, C.S. and Api, A.M. (2008) Fragrance material review
696 on borneol. *Food Chem Toxicol* **46**: S77-S80.

697 Chaw, S., Liu, Y., Wu, Y., Wang, H., Lin, C.I., Wu, C., Ke, H., Chang, L.,
698 Hsu, C., Yang, H., Sudianto, E., Hsu, M., Wu, K., Wang, L., Leebens-
699 Mack, J.H. and Tsai, I.J. (2019) Stout camphor tree genome fills gaps
700 in understanding of flowering plant genome evolution. *Nat Plants* **5**:
701 63-73.

702 Chen, N. (2004) Using RepeatMasker to Identify Repetitive Elements in
703 Genomic Sequences. *Curr Protoc Bioinformatics* **5**: 4-10.

704 Chen, L., Su, J., Li, L., Li, B. and Li, W. (2010b) A new source of natural
705 D-borneol and its characteristic. *J Med Plants Res* **5**: 15.
706 DeLano, W.L. (2002) Pymol: An open-source molecular graphics tool.
707 *CCP4 Newsletter on protein crystallography* **40**: 82-92.
708 Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M.,
709 Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P.
710 and Aiden, E.L. (2017) *De novo* assembly of the *Aedes aegypti*
711 genome using Hi-C yields chromosome-length scaffolds. *Science* **356**:
712 92.
713 Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P.,
714 Lander, E.S. and Aiden, E.L. (2016) Juicebox Provides a Visualization
715 System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**: 99-
716 101.
717 Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H.,
718 Lander, E.S. and Aiden, E.L. (2016) Juicer Provides a One-Click
719 System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*
720 **3**: 95-98.
721 Ellinghaus, D., Kurtz, S. and Willhoeft, U. (2008) LTRharvest, an efficient
722 and flexible software for de novo detection of LTR retrotransposons.
723 *BMC Bioinformatics* **9**: 18.
724 Ernst, J. and Bar-Joseph, Z. (2006) STEM: a tool for the analysis of short
725 time series gene expression data. *BMC Bioinformatics* **7**: 191.

726 Li, F.P., Gao, Y.H., Wu, B.Q., Cai, Q.P., Zhan, P.L., Yang, W.F., Shi, W.X.,
727 Li, X.H., Yang, Z.F., Tan, Q.Y., Luan, X., Zhang, G.Q. and Wang,
728 S.K.(2021) High-Quality de novo Genome Assembly of Huajingxian
729 74, a Receptor Parent of Single Segment Substitution Lines. *Rice Sci*
730 **28**: 109-113.
731 Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server:
732 interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29-
733 W37.
734 Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E.,
735 Gavin, O.L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L.,
736 Sonnhammer, E.L.L., Eddy, S.R. and Bateman, A. (2010) The Pfam
737 protein families database. *Nucleic Acids Res* **38**: D211-D222.
738 Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C.
739 and Smit, A.F. (2020) RepeatModeler2 for automated genomic
740 discovery of transposable element families. *PNAS* **117**: 9451.
741 Gouet, P., Courcelle, E., Stuart, D. and Métoz, F. (1999) ESPript: Analysis
742 of multiple sequence alignments in POSTScript. *Bioinformatics* **15**:
743 305-308.
744 Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White,
745 O., Buell, C.R. and Wortman, J.R. (2008) Automated eukaryotic gene
746 structure annotation using EVidenceModeler and the Program to
747 Assemble Spliced Alignments. *Genome Biol* **9**: R7.

748 Hu, J., Fan, J., Sun, Z. and Liu, S. (2019) NextPolish: A fast and efficient
749 genome polishing tool for long-read assembly. *Bioinformatics* **36**.
750 Huang, H., Lian, H., Boxiang, H.E., Wang, Y., Chen, G., Liang, D., Chen,
751 X., Luo, W., Lin, S. and Ling, F. (2019) Study on the Dynamic Change
752 of Essential Oil Content and Chemical Constituents in the Leaves of
753 *Cinnamomum burmannii* chvar. Borneol. *Forestry and Environmental*
754 *Sciences*.
755 Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence
756 Alignment Software Version 7: Improvements in Performance and
757 Usability. *Mol Biol Evol* **30**: 772-780.
758 Keilwagen, J., Hartung, F. and Grau, J. (2019) GeMoMa: Homology-Based
759 Gene Prediction Utilizing Intron Position Conservation and RNA-seq
760 Data. *Methods Mol Biol* **1962**: 161-177.
761 Kim, D., Langmead, B. and Salzberg, S. (2015) HISAT: A fast spliced
762 aligner with low memory requirements. *Nat Methods* **12(4)**: 357-362.
763 Kim, H., Lee, B., Won, E., Han, J., Hwang, D., Park, H.G. and Lee, J. (2015)
764 Identification of xenobiotic biodegradation and metabolism-related
765 genes in the copepod Tigriopus japonicus whole transcriptome
766 analysis. *Mar Genome* **24**:207-208.
767 Kumar, S., Stecher, G., Suleski, M. and Hedges, S.B. (2017) TimeTree: A
768 Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol*
769 *Evol* **34**:1812-1819.

770 Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M.,
771 Antonescu, C. and Salzberg, S.L. (2004) Versatile and open software
772 for comparing large genomes. *Genome Biol* **5**:R12.
773 Li, H. (2013) Aligning sequence reads, clone sequences and assembly
774 contigs with BWA-MEM. *eprint ArXiv* **1303**:3997.
775 Liao, Y., Smyth, G.K. and Shi, W. (2014) featureCounts: an efficient
776 general purpose program for assigning sequence reads to genomic
777 features. *Bioinformatics* **30**:923-930.
778 Lin, T., Xu, X., Ruan, J., Liu, S., Wu, S., Shao, X., Wang, X., Gan, L., Qin,
779 B., Yang, Y., Cheng, Z., Yang, S., Zhang, Z., Xiong, G., Huang, S., Yu,
780 H. and Li, J. (2018) Genome analysis of Taraxacum kok-saghyz Rodin
781 provides new insights into rubber biosynthesis. *Natl Sci Rev* **5**:78-87.
782 Liu, Z., Hernando, G., Suarez, D., Harnvanichvech, Y., Stephenson, M.J.,
783 Schranz, M.E., Nelson, D., Medema, M.H. and Osbour, A. (2020)
784 Drivers of metabolic diversification: how dynamic genomic
785 neighbourhoods generate new biosynthetic pathways in the
786 Brassicaceae. *New Phytol* **227**:1109-1123.
787 Love, M., Anders, S. and Huber, W. (2014) Differential analysis of count
788 data–the deseq2 package. *Genome Biol* **15**.
789 Lu, Y., Chen, X., Du, S., Wu, Q., Yao, Z. and Zhai, Y. (2010) The in situ
790 and in vivo study on enhancing effect of borneol in nasal absorption
791 of Geniposide in rats. *Arch Pharm Res* **33**:691-696.

792 Hiroyuki, O., Susumu, G., Kazushige, S., Wataru, F., Hidemasa, B. and
793 Minoru, K. (2000) KEGG: kyoto encyclopedia of genes and genomes.
794 *Nucleic Acids Res* **27**: 29-34.
795 Ma, R., Su, P., Guo, J., Jin, B., Ma, Q., Zhang, H., Chen, L., Mao, L., Tian,
796 M., Lai, C., Tang, J., Cui, G. and Huang, L. (2021) Bornyl
797 Diphosphate Synthase From *Cinnamomum burmanni* and Its
798 Application for (+)-Borneol Biosynthesis in Yeast. *Front Bioeng*
799 *Biotech* **9**:631863.
800 Nguyen, L., Schmidt, H.A., von Haeseler, A. and Minh. B.Q. (2015) IQ-
801 TREE: A Fast and Effective Stochastic Algorithm for Estimating
802 Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**:268-274.
803 Mao, L., Kawaide, H., Higuchi, T., Chen, M., Miyamoto, K., Hirata, Y.,
804 Kimura, H., Miyazaki, S., Teruya, M., Fujiwara, K., Tomita, K.,
805 Yamane, H., Hayashi, K., Nojiri, H., Jia, L., Qiu, J., Ye, C., Timko,
806 M.P., Fan, L. and Okada, K. (2020) Genomic evidence for convergent
807 evolution of gene clusters for momilactone biosynthesis in land plants.
808 *PNAS* **117**:12472.
809 Ou, S. and Jiang, N. (2018) LTR_retriever: A Highly Accurate and
810 Sensitive Program for Identification of Long Terminal Repeat
811 Retrotransposons. *Plant Physiol* **176**:1410-1422.
812 Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T. and
813 Salzberg, S.L. (2015) StringTie enables improved reconstruction of a

814 transcriptome from RNA-seq reads. *Nat Biotechnol* **33**:290-295.
815 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V. and
816 Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and
817 annotation completeness with single-copy orthologs. *Bioinformatics*
818 **31**:3210-3212.
819 Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J. and Holmes, I.H.
820 (2009) JBrowse: A next-generation genome browser. *Genome Res* **19**:
821 1630-1638.
822 Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., Xi, Z., Wang, X. and Liu,
823 J. (2021) WGDI: A user-friendly toolkit for evolutionary analyses of
824 whole-genome duplications and ancestral karyotypes.
825 Sun, X., Jiao, C., Schwaninger, H., Chao, C.T., Ma, Y., Duan, N., Khan, A.,
826 Ban, S., Xu, K., Cheng, L., Zhong, G. and Fei, Z. (2020) Phased
827 diploid genome assemblies and pan-genomes provide insights into the
828 genetic history of apple domestication. *Nat Genet* **52**:1423-1432.
829 Trott, O. and Olson, A.J. (2010) AutoDock Vina: Improving the speed and
830 accuracy of docking with a new scoring function, efficient
831 optimization, and multithreading. *J Comput Chem* **31**:455-461.
832 Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H.,
833 Gurtowski, J. and Schatz, M.C. (2017) GenomeScope: fast reference-
834 free genome profiling from short reads. *Binformatics* **33**:2202-2204.
835 Wang, S., Alseekh, S., Fernie, A.R. and Luo, J. (2019) The Structure and

836 Function of Major Plant Metabolite Modifications. *Mol Plant* **12**:899-
837 919.
838 Wolf, J.B., Bayer, T., Haubold, B, Schilhabel, M., Rosenstiel, P. and Tautz,
839 D. (2011) A new source of natural D-borneol and its characteristic. *J*
840 *Med Plants Res* **5**:162-175.
841 Wu, H., Zhu, L.F., and Li, Y.J. (1992) A STUDY ON INFRASPECIFIC
842 TYPES OF CHEMICAL CHARACTER OF CINNAMOMUM
843 BURMANNII. *J Integr Plant Biol* (in Chinese) **34**:302-308.
844 Xiao, C., Chen, Y., Xie, S., Chen, K., Wang, Y., Han, Y., Luo, F. and Xie,
845 Z. (2017) MECAT: fast mapping, error correction, and *de novo*
846 assembly for single-molecule sequencing reads. *Nat Methods*
847 **14**:1072-1074.
848 Xu, Z. and Wang, H. (2007) LTR-FINDER: An efficient tool for the
849 prediction of full-length LTR retrotransposons. *Nucleic Acids Res*
850 **35**:W265-W268.
851 Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood.
852 *Mol Biol Evol* **24**:1586-1591.
853 Zdobnov, E.M. and Apweiler, R. (2001) InterProScan-an integration
854 platform for the signature-recognition methods in InterPro.
855 *Bioinformatics* **17**:847-848.
856 Zuo, J., Grierson, D., Courtney, L., Wang, Y., Gao, L., Zhao, X., Zhu, B.,
857 Luo, Y., Wang, Q. and Giovannoni, J. (2020) Relationships between

genome methylation, levels of noncoding RNAs, mRNAs and metabolites in ripening tomato fruit. *The Plant Journal* **103**.