



Genome Sequencing of *Amomum tsao-ko* Provides Novel Insight Into Its Volatile Component Biosynthesis

Fenghui Sun¹, Chaochao Yan², Yunyun Lv^{2,3}, Zhonghui Pu^{1,4}, Zedong Liao¹, Wei Guo^{1,4*} and Min Dai^{1,4*}

OPEN ACCESS

Edited by:

Yunpeng Cao,
Wuhan Botanical Garden (CAS),
China

Reviewed by:

Mingcheng Wang,
Chengdu University, China
Chao Bian,
BGI Academy of Marine Sciences,
China
Tao Ma,
Sichuan University, China
Zhenxin Fan,
Sichuan University, China

*Correspondence:

Wei Guo
guowei@cmc.edu.cn
Min Dai
daimin1015@cmc.edu.cn

Specialty section:

This article was submitted to
Plant Bioinformatics,
a section of the journal
Frontiers in Plant Science

Received: 25 March 2022

Accepted: 09 May 2022

Published: 01 June 2022

Citation:

Sun F, Yan C, Lv Y, Pu Z, Liao Z,
Guo W and Dai M (2022) Genome
Sequencing of *Amomum tsao-ko*
Provides Novel Insight Into Its Volatile
Component Biosynthesis.
Front. Plant Sci. 13:904178.
doi: 10.3389/fpls.2022.904178

¹ School of Laboratory Medicine, Chengdu Medical College, Chengdu, China, ² Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China, ³ College of Life Science, Neijiang Normal University, Neijiang, China, ⁴ Sichuan Provincial Engineering Laboratory for Prevention and Control Technology of Veterinary Drug Residue in Animal-Origin Food, Chengdu Medical College, Chengdu, China

As an important economic and medicinal crop, *Amomum tsao-ko* is rich in volatile oils and widely used in food additives, essential oils, and traditional Chinese medicine. However, the lack of the genome remains a limiting factor for understanding its medicinal properties at the molecular level. Here, based on 288.72 Gb of PacBio long reads and 105.45 Gb of Illumina paired-end short reads, we assembled a draft genome for *A. tsao-ko* (2.70 Gb in size, contig N50 of 2.45 Mb). Approximately 90.07% of the predicted genes were annotated in public databases. Based on comparative genomic analysis, genes involved in secondary metabolite biosynthesis, flavonoid metabolism, and terpenoid biosynthesis showed significant expansion. Notably, the *DXS*, *GGPPS*, and *CYP450* genes, which participate in rate-limiting steps for terpenoid backbone biosynthesis and modification, may form the genetic basis for essential oil formation in *A. tsao-ko*. The assembled *A. tsao-ko* draft genome provides a valuable genetic resource for understanding the unique features of this plant and for further evolutionary and agronomic studies of Zingiberaceae species.

Keywords: genome, gene family expansion, terpenoid biosynthesis, flavonoid metabolic process, *Amomum tsao-ko*

INTRODUCTION

Amomum tsao-ko (Zingiberaceae) is a perennial herbaceous plant widely distributed in southwest China and Vietnam. As a traditional Chinese medicine, the dried fruits of *A. tsao-ko* are used to treat malaria, throat infections, abdominal pain, dyspepsia, nausea, stomach disorders, vomiting, and diarrhea (Tang et al., 2010). Clinical and animal trials indicate that *A. tsao-ko* exhibits a wide range of pharmacological activities, including antioxidant, cytotoxic, and antimicrobial activities (Li et al., 2014; Hong et al., 2015; Dai et al., 2016; Lin et al., 2021). The essential oil of *A. tsao-ko* and its polyphenol extract can modulate gut microbiota and alleviate hypercholesterolemia

(Liu et al., 2021). The ethanol extract of *A. tsao-ko* can improve dyslipidemia-related indices, including plasma levels of total cholesterol, low-density lipoprotein, high-density lipoprotein, and atherogenesis, in mice on high-carbohydrate diets (Park et al., 2021). It is also a common food additive and spice, which can develop food flavor while retaining medicinal effects.

Amomum tsao-ko-based essential oils include terpenoids, diarylheptanoids, bicyclic nonanes, and phenols, which may account for the plant's medicinal properties (Hong et al., 2015; Cui et al., 2017; Sim et al., 2019). The monoterpene alcohol geraniol is a widely used fragrance ingredient and one of the main components (13.69%) of *A. tsao-ko* essential oil (Lapczynski et al., 2008). Geraniol shows significant inhibitory effects against *Staphylococcus aureus*, a pathogen responsible for many infections (Long et al., 2020, 2022). Geraniol can also improve endothelial function in high-fat diet-fed mice by inhibiting oxidative stress (Wang et al., 2016). Eucalyptol (1,8-cineole), another component of *A. tsao-ko* essential oil, displays antioxidant, antibacterial, anti-inflammatory, and insecticidal activities (Cai et al., 2021). Furthermore, *A. tsao-ko* extract flavonoids show excellent antioxidative and antidiabetic activity (Zhang et al., 2022). While these studies have revealed the main medicinal properties and constituents of *A. tsao-ko*, the lack of a genome limits our understanding of the genomic and molecular basis of its volatile component biosynthesis.

Herein, we generated a draft genome assembly of *A. tsao-ko* using PacBio long reads and Illumina paired-end short reads. We constructed a genome-wide phylogeny of *A. tsao-ko* with eight other available plant genomes. Comparative genomic analysis indicated that gene families involved in the synthesis of terpenoids were expanded, which may provide clues for exploring the biosynthesis of volatile components in *A. tsao-ko*. Overall, this draft genome provides a valuable genetic resource for in-depth biological and evolutionary studies and for genetic improvement of *A. tsao-ko*.

MATERIALS AND METHODS

Sample Collection, Sequencing, and Data Qualification

We collected fresh leaves from an adult *A. tsao-ko* plant (Figure 1) growing in Guangxi Zhuang Autonomous Region, southern China. Total genomic DNA was extracted, and DNA quantification and quality testing were determined using NanoDrop 2000 spectrophotometry (Thermo Fisher Scientific), gel electrophoresis, and Qubit fluorometry (Invitrogen). For short-read sequencing, paired-end libraries with a 350-bp insert size were prepared following the manufacturer's instructions and then sequenced on the Illumina NovaSeq 6000 platform. Clean reads were obtained by removing contaminated reads from low-quality data. The PacBio single-molecule real-time (SMRT) bell library was constructed using a SMRTbell® Express Template Prep Kit 2.0 (Pacific Biosciences, PN 101-853-100). The library was prepared for sequencing on the PacBio Sequel II system (Pacific Biosciences, CA, United States). After adapter removal, we obtained subreads. A total of 105.45 Gb of raw

paired-end short reads and 288.72 Gb of PacBio subreads were generated, which were reduced by 0.09 and 0.12%, respectively, after trimming and quality control (Supplementary Table 1). Average subread length in the two PacBio libraries was 22,362 and 22,751 bp, with a mean insert length of 23,106 and 23,457 bp, respectively (Supplementary Table 2).

Fresh leaves were also prepared for RNA sequencing to aid in genome annotation. Total RNA was extracted using a QIAGEN® RNA Mini Kit following the manufacturer's protocols. RNA purity and integrity were assessed using the NanoPhotometer® spectrophotometer (IMPLEN, CA, United States) and RNA Nano 6000 Assay Kit and Bioanalyzer 2100 system (Agilent Technologies, CA, United States), respectively. The RNA sequencing library was constructed using a NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) following the manufacturer's instructions. RNA sequencing was performed on the Illumina NovaSeq 6000 platform. Low-quality reads were excluded using Trimmomatic v.0.36.23 (Bolger et al., 2014). After quality control, 7.42 Gb of clean data retained for genome annotation (Supplementary Table 1).

Genome Assembly and Quality Assessment

The 1C value of *A. tsao-ko* was measured using flow cytometry (Cytotflex, Bio-Rad, United States) with propidium iodide (PI) as the DNA stain and *Vigna radiata* as reference standard plant. The genome size of *V. radiata* is 579 Mb as described previously (Arumuganathan and Earle, 1991). Three biological repeats were performed and the mixture of two plants as internal standard. We then assembled the genome with PacBio long reads using mecat2 (Xiao et al., 2017) and polished the assembly with PacBio long reads and short pair-end reads using NextPolish v1.4.0.¹ The assembly quality was assessed using the Embryophyta gene sets in BUSCO v3.1.0 with genome mode and kmer-spectra analysis, referring to the previous studies (Simao et al., 2015; Mapleson et al., 2017; Yang et al., 2019; Yang F. S. et al., 2020; Wang et al., 2021). We also applied LTR assembly index (LAI) to evaluate the continuity of the assembly based on the ratio of whole LTR retrotransposons (LTR-RTs) (Ou et al., 2018). The genome quality is at a draft level when $0 < LAI \leq 10$, at reference level when $10 < LAI \leq 20$ and at gold level when $20 \leq LAI$ (Ou et al., 2018).

Genome Annotation and Analysis of Transposons

MITE-Hunter v1.0 was used to annotate the miniature inverted repeat transposable elements (MITEs) with default parameters (Han and Wessler, 2010). The LTR-retriever v2.82 pipeline, which combined results from LTRharvest (with parameter: -similar 90 -vic 10 -seed 20 -seqids yes -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1) and LTR-Finder v1.07 (-D 15000 -d 1000 -L 7000 -l 100 -p 20 -C -M 0.9), was used to identify long terminal repeat retrotransposons (LTR-RT) (Xu and Wang, 2007; Ellinghaus et al., 2008; Ou and Jiang, 2018). We then searched repetitive sequences in Repbase v20170127

¹<https://github.com/Nextomics/NextPolish>

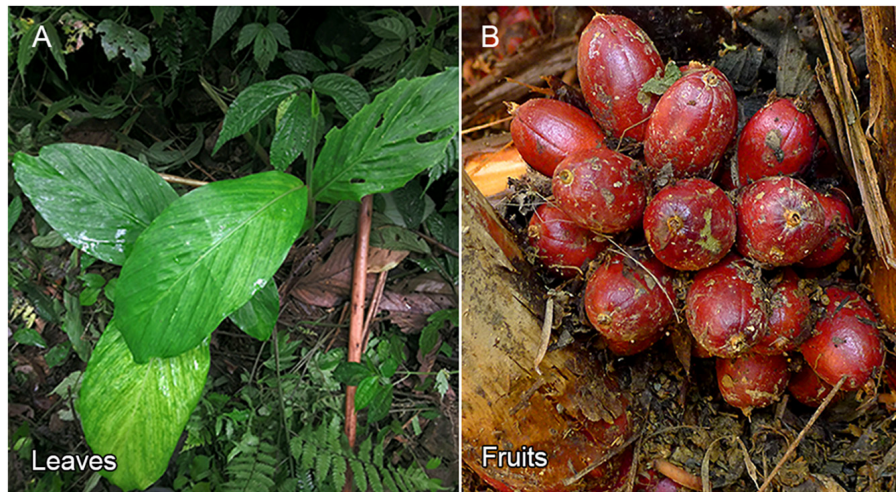


FIGURE 1 | Morphological features of *A. tsao-ko*. **(A)** The leaves of *A. tsao-ko*. **(B)** The fruits of *A. tsao-ko*.

using RepeatMasker v4.1.0 (Tarailo-Graovac and Chen, 2009) and constructed the repetitive sequence library by combining results from the MITE-Hunter and LTR-retriever pipelines. After masking the genome, *de novo* repeat annotation of *A. tsao-ko* was performed using RepeatModeler v2.0.1 (Flynn et al., 2020). Non-coding RNA was annotated against the sequence in the Rfam database and transfer RNA (tRNA) was predicted using tRNAscan-SE 2.0² (Chan et al., 2021).

A hybrid strategy of *de novo* gene prediction, homology-based prediction, and transcriptome alignment was applied for gene structural prediction using GeMoMa-1.6.1, Augustus v3.0.3, SNAP v6.0,³ GlimmerHMM v3.0.4, and GeneMark-ET v4.57 (Stanke and Waack, 2003; Majoros et al., 2004; Hoff et al., 2016; Keilwagen et al., 2016). The predicted results were combined using EVM⁴ and the untranslated region (UTR) and alternative splicing were predicted using PASA v2.0.1 (Haas et al., 2008). To determine the functional annotation of gene models, Diamond v0.9.31 (Buchfink et al., 2014) analysis with default parameters was performed against protein databases, including NR (non-redundant protein sequences in NCBI), SwissProt, and eggNOG, Gene Ontology (GO), describing molecular function (MF), cellular component (CC), and biology process (BP) terms, was annotated using Blast2GO (Conesa and Götz, 2008). Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis was used to annotate the pathway genes involved. The motifs and domains of each gene model were predicted using InterProScan v5.18-57.0 (Jones et al., 2014).

Phylogenetic Analyses

Referring to the genome study of a Zingiberaceae plant (Chakraborty et al., 2021), the orthologous groups among nine plant species, including *Ananas comosus* (GCF_001540865.1),

Asparagus officinalis (GCA_001876935.1), *Arabidopsis thaliana* (GCF_000001735.4), *Amborella trichopoda* (GCF_000471905.2), *Musa acuminata* (GCF_000313855.2), *Musa balbisiana* (GCA_004837865.1), *Oryza sativa* (GCF_001433935.1), *Sorghum bicolor* (GCF_000003195.3), and *A. tsao-ko*, were constructed using OrthoFinder v2.2.7 (Emms and Kelly, 2019). Single-copy genes from the nine species were extracted and the proteins for each gene were aligned. All alignments were combined to a supergene for each species to construct a phylogenetic tree using RAxML v8.2.12 (Stamatakis, 2014). Divergence time was estimated under the relaxed clock model using MCMCTree in PAML v4.9 (Yang, 2007). Three calibration points (the ancestors of *Asp. officinalis* and *M. acuminata*; *Ara. thaliana* and *Amb. trichopoda*; *O. sativa* and *S. bicolor*) for the divergence analysis were obtained from the TimeTree database (Kumar et al., 2017).

Analysis of Gene Family Expansion and Contraction

The results obtained from OrthoFinder v2.2.7 were used for gene family analysis. Genes that were unassigned (could not be clustered into any gene family) or found in only one species were considered species-specific. Gene family expansion and contraction analysis was performed using CAFE v4.2.1. A family-wide Viterbi *P*-value < 0.05 was defined as a significantly expanded or contracted gene family. Visualization used performed using python scripts.⁵

Functional and Pathway Enrichment Analysis

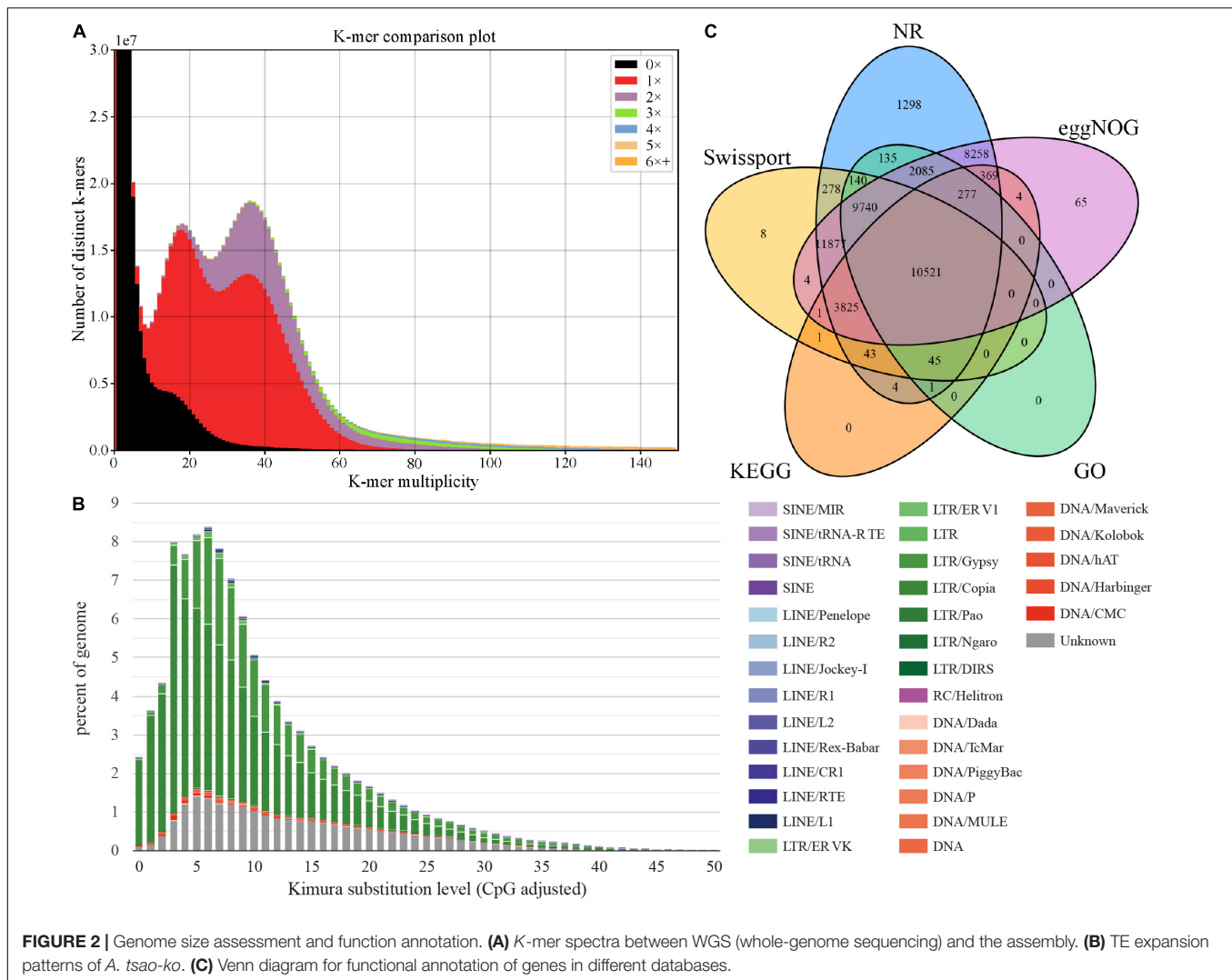
Enrichment analysis was performed to provide insights into the biological functions of species-specific genes and expanded genes families. GO and KEGG analyses were performed using the R package clusterProfiler v4.0 (Wu et al., 2021). The *A. tsao-ko* annotated results were set as background genes. Enriched terms

²<http://lowelab.ucsc.edu/tRNAscan-SE/>

³<http://snap.stanford.edu/snap/download.html>

⁴<https://sourceforge.net/projects/evidencemodeler/>

⁵https://github.com/hahnlab/cale_tutorial.git



with a corrected P -value < 0.05 were considered significantly over-represented.

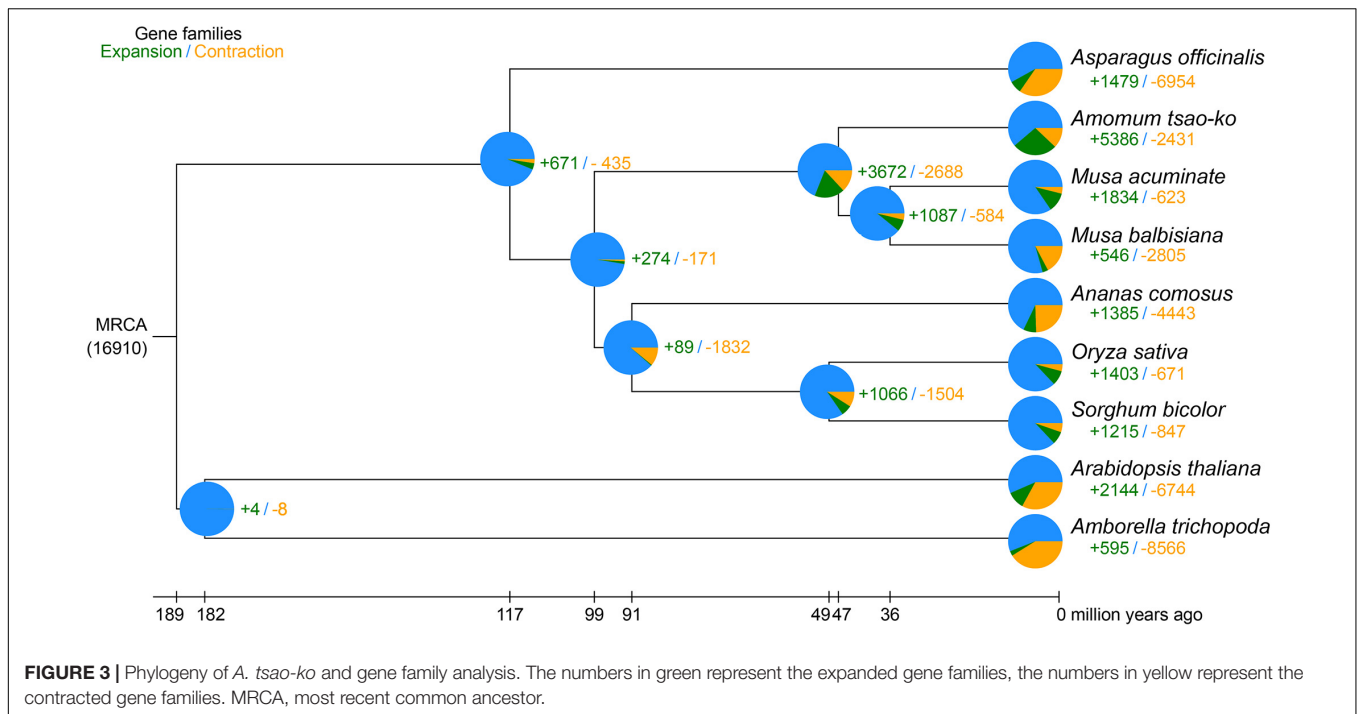
RESULTS AND DISCUSSION

De novo Assembly of *Amomum tsao-ko*

We estimate the genome size of *A. tsao-ko* with flow cytometry using *Vigna radiata* as reference standard and the results showed that the genome size of *A. tsao-ko* was approximately 3.17 Gb (**Supplementary Figure 1**). We used PacBio long reads to construct the primary assembly and used long reads and Illumina paired-end short reads to polish the assembly. The final *A. tsao-ko* assembly size was 2.70 Gb, with a contig N50 of 2.45 Mb. In comparison to other Zingiberaceae genomes, the *A. tsao-ko* genome had a higher contig N50 than turmeric (contig N50 = 0.1 Mb), but a lower contig N50 than ginger (contig N50: 4.68 Mb for haplotype 1 and 5.28 Mb for haplotype 0) (Chakraborty et al., 2021; Li et al., 2021). Average GC content in the *A. tsao-ko* genome was 41.07%, higher than that of ginger

(39.20%) and turmeric (38.75%). We evaluated assembly quality using BUSCO, resulting in 1,565 (97.0%) complete BUSCOs, including 1,117 (69.2%) single-copy BUSCOs, and 448 (27.8%) duplicated BUSCOs.

The *k*-mer Analysis Toolkit (KAT) can be used to assess errors, bias, and genome quality (Mapleson et al., 2017). We used KAT to estimate the assembly quality through pairwise comparison of *k*-mers present in both the input reads and assembly. As shown in **Figure 2A**, reads in black represent absence in the assembly, including incorrect and low-depth *k*-mers, accounting for a relatively small proportion. These suggest the current assembly covered most short reads *k*-mers content, with relatively high completeness (evaluation score 96.83%). We also observed multimodal spectra in the assembly, which may be influenced by heterozygous contents or by tetrapods of *A. tsao-ko* (Parthasarathy and Prasath, 2012). A previous study of 100 Archea, Bacteria, and Eukaryota species based on *k*-mer spectra indicates that species with multimodal spectra are consistent with tetrapods (Chor et al., 2009). Thus, KAT analysis of *A. tsao-ko* assembly indicated that the genome is complex and



cannot be explained by a simple probabilistic model, such as genome size estimation based on Poisson distribution of k -mer depth. GC-depth distribution showed two peaks at $\sim 20\times$ and $\sim 40\times$, also suggesting the complex of *A. tsao-ko* genome (high heterozygosity or polyploidy) (**Supplementary Figure 2**).

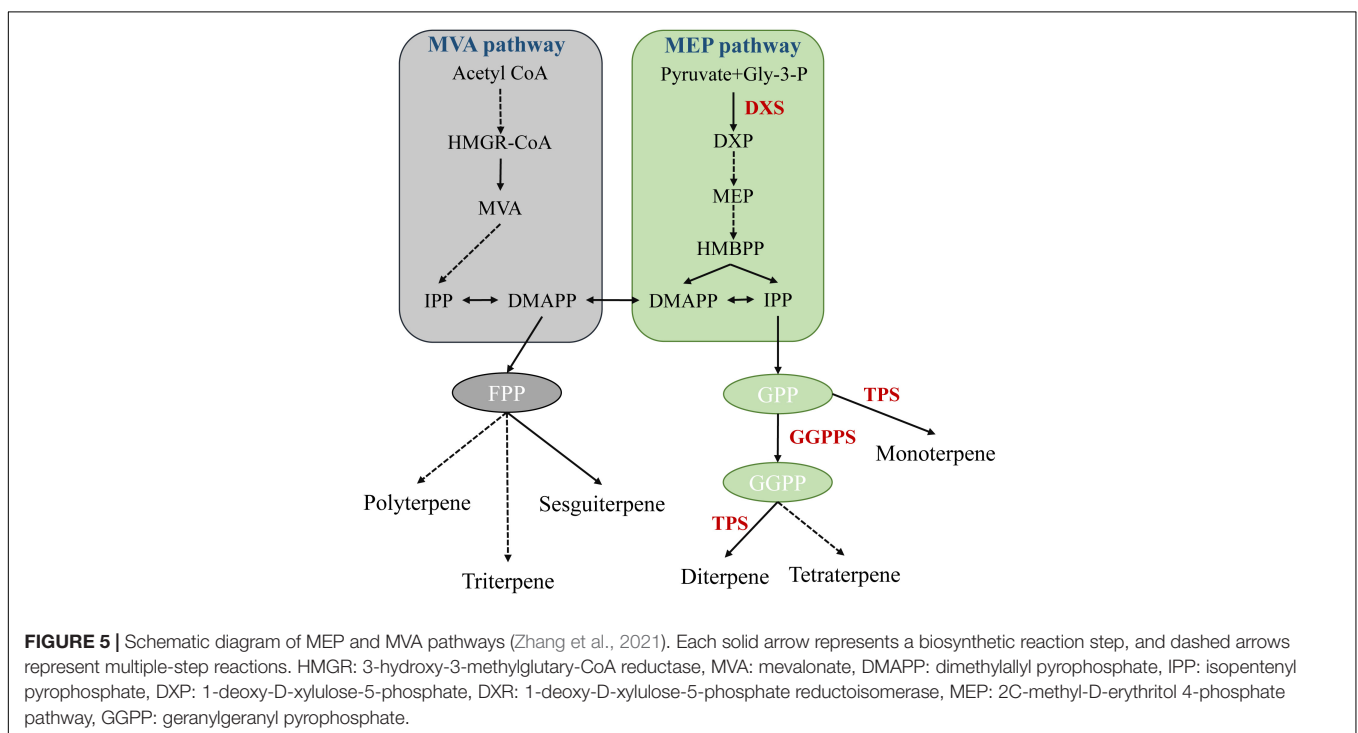
Containing appropriately 1,500 species, Zingiberaceae is one of the largest monocotyledonous families, producing valuable medicinal materials and spices. At present, however, only a few Zingiberaceae genomes have been reported, e.g., *Zingiber officinale* and *Curcuma longa* (Chakraborty et al., 2021; Li et al., 2021). Furthermore, within the *Amomum* genus, only a limited number of chloroplast genomes have been described (Zhang et al., 2019; Yang L. et al., 2020). The lack of whole-genome data has severely impeded our understanding of essential oil biosynthesis in *A. tsao-ko*. Thus, the genome reported in this study should serve as an important resource for further genetic improvement and for exploring the molecular basis of essential oil biosynthesis.

Repeat Identification and Gene Model Prediction

We annotated the repetitive sequences based on the *de novo* repeat sequence database of *A. tsao-ko* combined with Repbase v20170127. Results showed that 89.15% (2.41 Gb) of the *A. tsao-ko* genome contained repetitive sequences (**Supplementary Table 3**), much higher than that reported for other Zingiberaceae plants (e.g., $\sim 70\%$ for turmeric and $\sim 57\%$ for ginger) (Chakraborty et al., 2021; Li et al., 2021). Similar to turmeric [27.37% long terminal repeats (LTRs)], the LTR retrotransposons in the *A. tsao-ko* genome were also the most abundant transposable

elements, namely *LTR_Copia* and *LTR_Gypsy* (54.71%, **Figure 2B** and **Supplementary Table 3**). Simple tandem repeats accounted for a relatively low proportion (0.7%) of the *A. tsao-ko* genome. In addition, the LAI of the assembly was estimated as 17.85, suggesting a relatively high completeness.

In total, 54,379 protein-coding genes were annotated in the *A. tsao-ko* genome using the three strategies described above, and the number of genes predicted in *A. tsao-ko* (54,379), ginger ($\sim 39,000$), and turmeric (50,401) varied. Mean gene length was 5,613 bp and number of transcripts was 57,658, with 5.3 exons per transcript (**Supplementary Table 4**). Approximately 90.07% of the predicted genes were annotated in five public databases, including NR (89.92%), eggNOG (86.48%), SwissProt (67.09%), KEGG (27.75%), and GO (42.19%). The Venn diagram in **Figure 2C** shows that 10,521 (19.35%) protein-coding genes were simultaneously annotated in the five databases. GO annotation classified these genes into three main categories, i.e., biological process (e.g., oxidation-reduction process, proteolysis, and protein phosphorylation), cellular component (e.g., integral component of membrane, membrane, and nucleus), and molecular function (e.g., adenosine triphosphate (ATP) binding, metal ion binding, and zinc ion binding) (**Supplementary Figure 3**). In addition, non-coding RNAs were identified, resulting in 3,335 ribosomal RNAs (rRNAs), 234 microRNAs (miRNAs), and 1,417 tRNAs (**Supplementary Table 5**). Collinearity analysis found that the collinear duplicates occupied about 48% of all genes in *A. tsao-ko* genome. The K_s distribution displays a clear K_s peak (**Supplementary Figure 4**) at about 0.35 that should be resulted from a whole-genome duplication event, which leads to a tetraploid generation.



Single-Copy Orthologous Phylogeny

Nine plant species, i.e., *A. tsao-ko*, *Ana. comosus*, *Asp. officinalis*, *Ara. thaliana*, *Amb. trichopoda*, *M. acuminata*, *M. balbisiana*, *O. sativa*, and *S. bicolor*, were selected for orthologous group identification (Supplementary Table 6). In total, 1,288 single-copy orthologs shared among the species were extracted. We constructed a phylogenetic tree based on the 1,288 single-copy orthologs using RAxML, with *Ara. thaliana* and *Amb. trichopoda* set as the outgroups (Figure 3). The genome-wide phylogenetic positions of *A. tsao-ko* and selected species were supported by TIMETREE. Results showed that *M. acuminata*, *M. balbisiana*, and *A. tsao-ko* belonged to Zingiberales, and

shared the same phylogenetic clade. Furthermore, *A. tsao-ko* separated from Musaceae approximately 30~63 million years ago (MYA) and *Asp. officinalis*, as a member of Asparagales, showed early divergence among the monocotyledons.

Analysis of Gene Families and Genes Involved in Flavonoid Metabolism

To investigate the genetic basis of essential oil biosynthesis, we performed gene family expansion and contraction analysis of *A. tsao-ko* in comparison to the other eight species selected. Notably, 5,386 gene families showed significant expansion and 2,431 gene families showed significant contraction in

A. tsao-ko (family-wide Viterbi P -value < 0.05, **Figure 3**). The expanded gene families were subjected to functional enrichment analysis (P -adjust cutoff of 0.05). The top 10 most significantly enriched terms included cellular macromolecule metabolic process (GO:0044260), endonuclease activity (GO:0004519), and peptidase activity (GO:0008233) (**Supplementary Table 7**). In addition, multiple biosynthetic and metabolic process-related terms were significantly enriched, including secondary metabolite biosynthetic process (GO:0044550), S-adenosylmethionine biosynthetic process (GO:0006556), one-carbon metabolic process (GO:0006730), and flavonoid metabolic process (GO:0009812) (**Figure 4A**). The essential oil of *A. tsao-ko* is a secondary metabolite with strong biological activity and medicinal value and plays an important role in plant defense against disease, insects, and competition (Qin et al., 2021). Thus, the significant expansion of genes associated with secondary metabolism suggests enhancement of related functions.

Previous studies have suggested that *A. tsao-ko* shows potential as a novel drug for the treatment of type 2 diabetes due to the excellent antioxidative and antidiabetic activity of its flavonoids (Fang et al., 2019; Zhang et al., 2022). Here, flavonoid metabolic processes were significantly enriched by the *A. tsao-ko* expanded genes, including *UGT71K1*, *RGGA*, and *CZOG2*. Among these genes, *UGT71K1* encodes a protein with chalcone and flavonol 2'-O-glycosyltransferase activity, as well as, glycosyltransferase activity toward quercetin isoliquiritigenin and butein (Gosch et al., 2010). Flavanols, as major components of flavonoids in *A. tsao-ko* extract, show antidiabetic potency (Fang et al., 2019; He et al., 2020, 2021). In addition, *UGT71K1* can convert phloretin to phlorizin, a potent antioxidant with antidiabetic effects that competitively inhibits sodium-glucose symporters (Gosch et al., 2010).

Pathway Enrichment Analysis and Genetic Basis of Terpenoid Biosynthesis

We performed KEGG pathway enrichment analysis of the expanded genes. Results showed that the expanded genes in *A. tsao-ko* were significantly enriched in biosynthesis- [i.e., monoterpene biosynthesis (ko00902), terpenoid backbone biosynthesis (ko00900), tropane, piperidine and pyridine alkaloid biosynthesis (ko00960), and stilbenoid, diarylheptanoid and gingerol biosynthesis (ko00945)], metabolism- [e.g., metabolism of xenobiotics by cytochrome P450 (ko00980) and drug metabolism-cytochrome P450 (ko00982)], and immune-related pathways [e.g., plant-pathogen interaction (ko04626) and melanogenesis (ko04916)] (**Figure 4B** and **Supplementary Table 8**).

Terpenoids, such as geraniol and eucalyptol, are the main components of *A. tsao-ko* essential oil, and show antioxidant, antidiabetic, antibacterial, anti-inflammatory, and insecticidal activities (Lapczynski et al., 2008; Dai et al., 2016; Wang et al., 2016; Cai et al., 2021). Biosynthesis of terpenoids in plants is a complex process, involving backbone biosynthesis and terpenoid synthesis and modification. In nature, mevalonate (MVA) and 2C-methyl-d-erythritol 4-phosphate (MEP), located in the cytoplasm and plastids, respectively, are two major

pathways of terpenoid biosynthesis (**Figure 5**; Lei et al., 2021). We found that 1-deoxy-D-xylulose-5-phosphate synthase (including *DXS*, *DXS1*, and *DXS2*) and geranylgeranyl diphosphate synthase (*GGPPS*) genes were significantly expanded and enriched in the terpenoid backbone biosynthesis pathway. *DXS* can catalyze the condensation of pyruvate and d-glyceraldehyde 3-phosphate (GAP) to produce 1-deoxy-D-xylulose 5-phosphate (DXP), the first rate-limiting reaction of the MEP pathway (Hahn et al., 2001; Battistini et al., 2016). *GGPP* serves as a key precursor substrate of volatile and non-volatile terpenoids (Beck et al., 2013). *GGPPS* encodes an important enzyme involved in the synthesis of volatile and non-volatile terpenoids, constituting a key node that regulates carbon flow in the isoprenoid pathway (Zhang et al., 2021). Furthermore, based on Café pipeline analysis, terpene synthase (TPS) genes, including *TPS2*, *TPS4*, and *TPS10*, were expanded in *A. tsao-ko*, and significantly over-represented in the monoterpene biosynthesis pathway. TPSs can harness specific prenyl precursors to produce hemiterpenoids, monoterpene, sesquiterpenoids, diterpenoids, triterpenoids, and tetraterpenoids (Lei et al., 2021). Cytochrome P450 (CYP450) enzymes play critical roles in terpenoid skeleton modification and structural diversity (Zheng et al., 2019), and we found that related pathways were also over-represented in various expanded genes, such as *GSTU6*, *GSTUF*, and *GSTF1*. Overall, the expansion of genes encoding key rate-limiting enzymes in terpenoid synthesis and modification-related pathways may facilitate the synthesis of terpenoids, highlighting the biological activity and medicinal properties of *A. tsao-ko*.

CONCLUSION AND FUTURE PERSPECTIVES

In this study, we assembled a draft genome of *A. tsao-ko*, which should provide valuable insights into the evolutionary history of Zingiberaceae. We further identified candidate genes involved in the biosynthesis of terpenoids, flavonoids, and other secondary metabolites, including several genes encoding key rate-limiting enzymes of the biosynthetic pathway. These results provide a genetic basis for the formation of main terpenoids and other secondary metabolites of *A. tsao-ko*, which is of great advantage for the manipulation of related enzymes and improvement of breeding of this important medicinal plant. However, given the complexity of the *A. tsao-ko* genome, further studies are needed.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb), accession number CNP0002802.

AUTHOR CONTRIBUTIONS

WG and MD conceived the study and worked on the approval of the manuscript. FS, WG, and MD prepared the initial

manuscript draft. FS and ZL performed data analyses. CY finished evolutionary analysis. YL assembled the genome. ZP collected experimental samples and modified the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a grant from the Fund of the National Natural Science Foundation of China (Nos. 82102442, 31900307, and 31970137), the Sichuan Science and Technology Program (Nos. 2020JDRC0071, 2020YJ0401, and 2021YJ0158), the project funded by China Postdoctoral Science Foundation (No. 2021M703134), the China Scholarship

Council (No. 201908515091), and the Development and Regeneration Key Laboratory of Sichuan Province, Chengdu Medical College (No. SYS19-08).

ACKNOWLEDGMENTS

We would like to thank Wei Wu for analyzing the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.904178/full#supplementary-material>

REFERENCES

- Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species nuclear DNA content material and methods. *Plant Mol. Biol. Report.* 9, 208–218.
- Battistini, M. R., Shoji, C., Handa, S., Breydo, L., and Merkle, D. J. (2016). Mechanistic binding insights for 1-deoxy-d-Xylulose-5-Phosphate synthase, the enzyme catalyzing the first reaction of isoprenoid biosynthesis in the malaria-causing protists, *Plasmodium falciparum* and *Plasmodium vivax*. *Protein Expr. Purif.* 120, 16–27. doi: 10.1016/j.pep.2015.12.003
- Beck, G., Coman, D., Herren, E., Ruiz-Sola, M. Á., Rodríguez-Concepción, M., Grussem, W., et al. (2013). Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* 82, 393–416. doi: 10.1007/s11103-013-0070-z
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Cai, Z. M., Peng, J. Q., Chen, Y., Tao, L., Zhang, Y. Y., Fu, L. Y., et al. (2021). 1,8-Cineole: a review of source, biological activities, and application. *J. Asian Nat. Prod. Res.* 23, 938–954. doi: 10.1080/10286020.2020.1839432
- Chakraborty, A., Mahajan, S., Jaiswal, S. K., and Sharma, V. K. (2021). Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Commun. Biol.* 4, 1–12. doi: 10.1038/s42003-021-02720-y
- Chan, P. P., Lin, B. Y., Mak, A. J., and Lowe, T. M. (2021). TRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* 49, 9077–9096. doi: 10.1093/nar/gkab688
- Chor, B., Horn, D., Goldman, N., Levy, Y., and Massingham, T. (2009). Genomic DNA k-mer spectra: models and modalities. *Genome Biol.* 10:R108. doi: 10.1186/gb-2009-10-10-r108
- Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832
- Cui, Q., Wang, L. T., Liu, J. Z., Wang, H. M., Guo, N., Gu, C. B., et al. (2017). Rapid extraction of *Amomum tsao-ko* essential oil and determination of its chemical composition, antioxidant and antimicrobial activities. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1061–1062, 364–371. doi: 10.1016/j.jchromb.2017.08.001
- Dai, M., Peng, C., Peng, F., Xie, C., Wang, P., and Sun, F. (2016). Anti-Trichomonas vaginalis properties of the oil of *Amomum tsao-ko* and its major component, geraniol. *Pharm. Biol.* 54, 445–450. doi: 10.3109/13880209.2015.1044617
- Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18. doi: 10.1186/1471-2105-9-18
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 1–14. doi: 10.1186/s13059-019-1832-y
- Fang, J. Y., Lin, C. H., Huang, T. H., and Chuang, S. Y. (2019). In vivo rodent models of type 2 diabetes and their usefulness for evaluating flavonoid bioactivity. *Nutrients* 11:530. doi: 10.3390/nu11030530
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Gosch, C., Halbwirth, H., Schneider, B., Hölscher, D., and Stich, K. (2010). Cloning and heterologous expression of glycosyltransferases from *Malus x domestica* and *Pyrus communis*, which convert phloretin to phloretin 2'-O-glucoside (phloridzin). *Plant Sci.* 178, 299–306. doi: 10.1016/j.plantsci.2009.12.009
- Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9, 1–22. doi: 10.1186/gb-2008-9-1-r7
- Hahn, F. M., Eubanks, L. M., Testa, C. A., Blagg, B. S. J., Baker, J. A., and Poulter, C. D. (2001). 1-deoxy-D-xylulose 5-phosphate synthase, the gene product of open reading frame (ORF) 2816 and ORF 2895 in *Rhodobacter capsulatus*. *J. Bacteriol.* 183, 1–11. doi: 10.1128/JB.183.1.1-11.2001
- Han, Y., and Wessler, S. R. (2010). MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* 38, 1–8. doi: 10.1093/nar/gkq862
- He, X. F., Chen, J. J., Huang, X. Y., Hu, J., Zhang, X. K., Guo, Y. Q., et al. (2021). The antidiabetic potency of *Amomum tsao-ko* and its active flavanols, as PTP1B selective and α -glucosidase dual inhibitors. *Ind. Crops Prod.* 160:112908. doi: 10.1016/j.indcrop.2020.112908
- He, X. F., Chen, J. J., Li, T. Z., Zhang, X. K., Guo, Y. Q., Zhang, X. M., et al. (2020). Nineteen new flavanol-fatty alcohol hybrids with α -glucosidase and PTP1B dual inhibition: one unusual type of antidiabetic constituent from *Amomum tsao-ko*. *J. Agric. Food Chem.* 68, 11434–11448. doi: 10.1021/acs.jafc.0c04615
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., and Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32, 767–769. doi: 10.1093/bioinformatics/btv661
- Hong, S. S., Lee, J. H., Choi, Y. H., Jeong, W., Ahn, E. K., Lym, S. H., et al. (2015). Amotsaokonal A-C, benzaldehyde and cycloterpenal from *Amomum tsao-ko*. *Tetrahedron Lett.* 56, 6681–6684. doi: 10.1016/j.tetlet.2015.10.045
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* 44, 1–11. doi: 10.1093/nar/gkw092
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116
- Lapczynski, A., Bhatia, S. P., Foxenberg, R. J., Letizia, C. S., and Api, A. M. (2008). Fragrance material review on geraniol. *Food Chem. Toxicol.* 46, 160–170.

- Lei, D., Qiu, Z., Qiao, J., and Zhao, G. R. (2021). Plasticity engineering of plant monoterpene synthases and application for microbial production of monoterpenoids. *Biotechnol. Biofuels* 14, 1–15.
- Li, B., Choi, H. J., Lee, D. S., Oh, H., Kim, Y. C., Moon, J. Y., et al. (2014). *Amomum tsao-ko* suppresses lipopolysaccharide-induced inflammatory responses in RAW264.7 macrophages via Nrf2-dependent heme oxygenase-1 expression. *Am. J. Chin. Med.* 42, 1229–1244. doi: 10.1142/S0192415X14500773
- Li, H. L., Wu, L., Dong, Z., Jiang, Y., Jiang, S., Xing, H., et al. (2021). Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. *Hortic. Res.* 8:530. doi: 10.1038/s41438-021-00627-7
- Lin, L., Long, N., Qiu, M., Liu, Y., Sun, F., and Dai, M. (2021). The inhibitory efficiencies of geraniol as an anti-inflammatory, antioxidant, and antibacterial, natural agent against methicillin-resistant staphylococcus aureus infection in vivo. *Infect. Drug Resist.* 14, 2991–3000. doi: 10.2147/IDR.S318989
- Liu, L., Zhao, Y., Ming, J., Chen, J., Zhao, G., Chen, Z.-Y., et al. (2021). Polyphenol extract and essential oil of *Amomum tsao-ko* equally alleviate hypercholesterolemia and modulate gut microbiota. *Food Funct.* 12, 12008–12021. doi: 10.1039/d1fo03082e
- Long, N., Tang, H., Lin, L., Sun, F., Peng, C., and Dai, M. (2020). Activity of *Amomum tsao-ko* fruits essential oil against methicillin-resistant *Staphylococcus aureus* in vivo. *HSOA J. Altern. Complement. Integr. Med.* 6:126. doi: 10.24966/acim-7562/100126
- Long, N., Zhang, Y., Qiu, M., Deng, J., Sun, F., and Dai, M. (2022). Dynamic changes of inflammatory response and oxidative stress induced by methicillin-resistant *Staphylococcus aureus* in mice. *Eur. J. Clin. Microbiol. Infect. Dis.* 41, 79–86. doi: 10.1007/s10096-021-04349-5
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315
- Mapleson, D., Accinelli, G. G., Kettleborough, G., Wright, J., and Clavijo, B. J. (2017). KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33, 574–576. doi: 10.1093/bioinformatics/btw663
- Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res.* 46, e126. doi: 10.1093/nar/gky730
- Ou, S., and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310
- Park, J. H., Ahn, E. K., Hwang, M. H., Park, Y. J., Cho, Y. R., Ko, H. J., et al. (2021). Improvement of obesity and dyslipidemic activity of *Amomum tsao-ko* in c57bl/6 mice fed a high-carbohydrate diet. *Molecules* 26:1638. doi: 10.3390/molecules26061638
- Parthasarathy, V. A., and Prasath, D. (2012). Cardamom: possibilities for sustainable cultivation in the East Usambaras, Tanzania. *Handb. Herbs Spices Sec. Ed.* 1, 131–170. doi: 10.1533/9780857095671.131
- Qin, H., Wang, Y., Yang, W., Yang, S., and Zhang, J. (2021). Comparison of metabolites and variety authentication of *Amomum tsao-ko* and *Amomum paratsao-ko* using GC-MS and NIR spectroscopy. *Sci. Rep.* 11, 1–12. doi: 10.1038/s41598-021-94741-0
- Sim, S., Tan, S. K., Kohlenberg, B., and Braun, N. A. (2019). *Amomum tsao-ko*-Chinese black cardamom: detailed oil composition and comparison with two other cardamom species. *Nat. Prod. Commun.* 14, 1–12. doi: 10.1177/1934578X19857675
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO online supplementary information: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. doi: 10.1093/bioinformatics/btv351
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, 215–225. doi: 10.1093/bioinformatics/btg1080
- Tang, W., Eisenbrand, G., and Ajjan, R. (2010). *Handbook of Chinese Medicinal Plants: Chemistry, Pharmacology, Toxicology*. Weinheim: Wiley-VCH Verlag GmbH.
- Tarailo-Graovac, M., and Chen, N. (2009). Using repeatmasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* Chapter 4:Unit4.10. doi: 10.1002/0471250953.bi0410s25
- Wang, M., Tong, S., Ma, T., Xi, Z., and Liu, J. (2021). Chromosome-level genome assembly of *Sichuan pepper* provides insights into apomixis, drought tolerance, and alkaloid biosynthesis. *Mol. Ecol. Resour.* 21, 2533–2545. doi: 10.1111/1755-0998.13449
- Wang, X., Zhao, S., Su, M., Sun, L., Zhang, S., Wang, D., et al. (2016). Geraniol improves endothelial function by inhibiting NOX-2 derived oxidative stress in high fat diet fed mice. *Biochem. Biophys. Res. Commun.* 474, 182–187. doi: 10.1016/j.bbrc.2016.04.097
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* 2:100141. doi: 10.1016/j.xinn.2021.100141
- Xiao, C. L., Chen, Y., Xie, S. Q., Chen, K. N., Wang, Y., Han, Y., et al. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14, 1072–1074. doi: 10.1038/nmeth.4432
- Xu, Z., and Wang, H. (2007). LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35, 265–268. doi: 10.1093/nar/gkm286
- Yang, F. S., Nie, S., Liu, H., Shi, T. L., Tian, X. C., Zhou, S. S., et al. (2020). Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat. Commun.* 11, 1–13. doi: 10.1038/s41467-020-18771-4
- Yang, L., Feng, C., Cai, M., Chen, J., and Ding, P. (2020). Complete chloroplast genome sequence of *Amomum villosum* and comparative analysis with other Zingiberaceae plants. *Chin. Herb. Med.* 12, 375–383. doi: 10.1016/j.chmed.2020.05.008
- Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., et al. (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* 51, 1052–1059. doi: 10.1038/s41588-019-0427-6
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/m sm088
- Zhang, C. G., Liu, H. H., Zong, Y. X., Tu, Z. H., and Li, H. G. (2021). Isolation, expression, and functional analysis of the geranylgeranyl pyrophosphate synthase (GGPPS) gene from *Liriodendron tulipifera*. *Plant Physiol. Biochem.* 166, 700–711. doi: 10.1016/j.plaphy.2021.06.052
- Zhang, X. F., Tang, Y. J., Guan, X. X., Lu, X., Li, J., Chen, X. L., et al. (2022). Flavonoid constituents of *Amomum tsao-ko* Crevost et Lemarie and their antioxidant and antidiabetic effects in diabetic rats- in vitro and in vivo studies. *Food Funct.* 13, 437–450. doi: 10.1039/d1fo02974f
- Zhang, Y.-M., Yang, C.-W., Liu, Y.-Y., Yang, Y.-W., Liu, X.-L., and Li, G.-D. (2019). Complete chloroplast genome sequences of two *Amomum* species (Zingiberaceae). *Mitochondrial DNA B Resour.* 4, 3795–3796. doi: 10.1080/23802359.2019.1682951
- Zheng, X., Li, P., and Lu, X. (2019). Research advances in cytochrome P450-catalysed pharmaceutical terpenoid biosynthesis in plants. *J. Exp. Bot.* 70, 4619–4630. doi: 10.1093/jxb/erz203

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Sun, Yan, Lv, Pu, Liao, Guo and Dai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.