Comparing genomes of Fructus Amomi-producing species reveals genetic basis of volatile terpenoid divergence

Peng Yang[1,2,5], Xu-Yi Ling[1,5], Xiao-Fan Zhou[3,5], Yuan-Xia Chen[1], Tian-Tian Wang[1], Xiao-Jing Lin[1], Yuan-Yuan Zhao[1], Yu-Shi Ye[4], Lin-Xuan Huang[1], Ye-Wen Sun[1], Yu-Xin Qi[2], Dong-Ming Ma[1], Ruo-Ting Zhan[1,*], Xue-Shuang Huang[2,*], Jin-Fen Yang[1,*,†]

1. Key Laboratory of Chinese Medicinal Resource from Lingnan (Ministry of Education), School of Pharmaceutical Science, Guangzhou University of Chinese Medicine, Guangzhou 510006, China
2. Hunan Provincial Key Laboratory for Synthetic Biology of Traditional Chinese Medicine, School of Pharmaceutical Sciences, Hunan University of Medicine, Huaihua 418000, China
3. Guangdong Province Key Laboratory of Microbial Signals and Disease Control, Integrative Microbiology Research Centre, South China Agricultural University, Guangzhou 510642, China
4. South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
5. These authors contributed equally to this article

*Corresponding authors: yangjf@gzucm.edu.cn (J.F.Y), xueshuanghuang@126.com (X.S.H), zhanrt@gzucm.edu.cn (R.T.Z)

†Senior author

Short title: Dissection of volatile terpenoids in *Fructus Amomi*

One-sentence summary:

Functional differentiation of bornyl diphosphate synthase and positive regulation of GCN4-motif contribute to the difference of volatile terpenoids in *W. longiligularis* and *W. villosa*.

The author responsible for distribution of materials integral to the findings presented in

this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/general-instructions) is: Jin-Fen Yang (yangjf@gzucm.edu.cn).

Abstract

Wurfbainia longiligularis and *Wurfbainia villosa* are both rich in volatile terpenoids and are two primary plant sources of Fructus Amomi used for curing gastrointestinal diseases. Metabolomic profiling has demonstrated that bornyl diphosphate (BPP)-related terpenoids are more abundant in the *W. villosa* seeds and have a wider tissue distribution in *W. longiligularis*. To explore the genetic mechanisms underlying the volatile terpenoids divergence, a high-quality chromosome-level genome of *W. longiligularis* (2.29 Gb, contig N50 of 80.39 Mb) was assembled. Functional characterization of 17 terpene synthases (WlTPSs) revealed that WlBPPS, along with WlTPS 24/26/28 with bornyl diphosphate synthase (BPPS) activity, contribute to the wider tissue distribution of BPP-related terpenoids in *W. longiligularis* compared to *W. villosa*. Furthermore, transgenic *Nicotiana tabacum* showed that the GCN4-motif element positively regulates seed expression of *WvBPPS* and thus promotes the enrichment of BPP-related terpenoids in *W. villosa* seeds. Systematic identification and analysis of candidate TPS in 29 monocot plants from 16 families indicated that substantial expansion of TPS-a and TPS-b subfamily genes in *Zingiberaceae* may have driven increased diversity and production of volatile terpenoids. Evolutionary analysis and functional identification of BPPS genes showed that BPP-related terpenoids may be distributed only in the *Zingiberaceae* of

monocot plants. This research provides valuable genomic resources for breeding and improving *Fructus Amomi* with medicinal and edible value and sheds light on the evolution of terpenoid biosynthesis in *Zingiberaceae*.

Keywords: *Wurfbainia longiligularis*, bornyl diphosphate synthase, terpene synthase, volatile terpenoids, comparative genomics

Introduction

Borneol, bornyl acetate, and camphor are aromatic and medicinal compounds found in many plants and are widely utilized in the pharmaceutical, chemical, and food industries (Mei et al., 2023; Zhao et al., 2023; Bahramikia et al., 2022). Bornyl diphosphate synthase (BPPS) is a terpene cyclase in the terpene synthase (TPS) gene family that catalyzes the production of bornyl diphosphate (BPP), the precursor of borneol, bornyl acetate and camphor biosynthesis, from geranyl diphosphate (GPP) (Whittington et al., 2002). Its activity and expression levels directly impact the content of BPP-related terpenoids in plants. For example, *Salvia officinalis* BPPS produces 75% BPP and 25% monoterpenoids, while *Lavandula angustifolia* BPPS generates 30% BPP and 70% monoterpenoids (Adal et al., 2023; Matthew et al., 2017). To date, the biosynthesis of volatile terpenoids in *Wurfbainia villosa* has been well-characterized, and WvBPPS is mainly responsible for the synthesis of BPP-related terpenoids in the seeds (Wang et al., 2018; Yang et al., 2022). There are perceptible differences in the BPP-related terpenoid content and tissue distribution between *Wurfbainia longiligularis* and *W. villosa*, which are both plant sources of Fructus Amomi (Doh et al., 2020). However, the biosynthesis of volatile terpenoids in *W. longiligularis* and the genetic underpinnings of the BPP-related terpenoid

divergence between the two species are currently unknown. Chinese medicine often involves numerous plant sources and diverse active ingredients, such as Gancao (*Glycyrrhiza uralensis*) and Danshen (*Salvia miltiorrhiza*). Applying modern omics techniques to explore the genetic basis of such differences can provide further insights into improving the quality of medicinal materials (Ma et al., 2021; Zhong et al., 2022). Therefore, a comparative genomics analysis of *W. longiligularis* and *W. villosa* can provide further insights into BPP-related terpenoid biosynthesis.

Volatile terpenoids are the largest group of volatile organic compounds released by plants, including isoprenoids, monoterpenoids and sesquiterpenoids, which play a crucial role in plant-to-plant signaling and disease management (Jia et al., 2022; Bao et al., 2023; Rosenkranz et al., 2021). Despite their structural diversity, all volatile terpenoids are generated by TPSs, especially TPS-a and TPS-b subfamilies, from GPP and farnesyl diphosphate (FPP) to produce monoterpenoids and sesquiterpenoids, respectively (Nagegowda et al., 2020; Guo et al., 2021). *Zingiberaceae* plants, particularly medicinal ones, have been reported to have the most abundant volatile terpenoids among monocot plants (Barbosa et al., 2017; Tunnisa et al., 2022; Kulyal et al., 2021; Peng et al., 2022; Ivanovic et al., 2021). So far, the biosynthesis of volatile terpenoids has only been studied comprehensively in *W. villosa* of *Zingiberaceae* plants. Moreover, only BPP-related terpenoids have been found in *Zingiberaceae* in monocot plants, and the evolution and function of BPPS genes in this group of plants remain poorly understood (Yang et al., 2022). In recent years, a growing number of researchers have focused on multi-omics studies of *Zingiberaceae* plants to publish multi-plant chromosome-level genomes,

including Ginger (*Zingiber officinale*), *W. villosa*, *Lanxangia tsao-ko*, *Curcuma alismatifolia*, and *Curcuma longa* (Li et al., 2021; Yang et al., 2022; Liao et al., 2022; Yin et al., 2022; Li et al., 2022). These multi-omics data will further advance research on the evolution of the TPS gene family and the biosynthesis of volatile terpenoids in *Zingiberaceae* plants.

Here, we generated a high-quality chromosome-level genome of *W. longiligularis* by combining PacBio HiFi data and Hi-C sequencing technologies. Comparative genomics revealed the contribution of a *Zingiberaceae*-wide whole-genome duplication (WGD) event and a recent burst of long terminal repeats (LTRs) in the genome evolution of *W. longiligularis*. Functional characterization of 17 WlTPSs and discovery that the biosynthesis and tissue distribution of BPP-related terpenoids were driven by the tandem duplication of *WlBPPS*. In addition, the GCN4-motif element that could positively regulate the seed expression of *WvBPPS* was confirmed by using transgenic *N. tabacum*. Synteny analysis and functional identification of BPPS genes provided insights into the metabolic evolution of BPP-related terpenoids in *Zingiberaceae* plants. By comparing the catalytic activity of BPPS from different plant sources and performing site-directed mutagenesis, we identified seven key amino acid residues that affect WvBPPS activity. In summary, our comprehensive approach involving genomics, transcriptomics, metabolomics, and biochemistry has shed light on volatile terpenoid biosynthesis and BPPS evolution in *W. longiligularis*, and will hopefully facilitate varietal improvement and molecular breeding of *Fructus Amomi*.

Results

Distribution of volatile terpenoids in *W. longiligularis* and *W. villosa*

To compare volatile terpenoids in *W. longiligularis* and *W. villosa* (Figures 1A and 1B), we performed qualitative and quantitative analyses of seven different tissues using the GC-MS method. A total of 40 and 41 volatile terpenoids were identified in *W. longiligularis* and *W. villosa*, respectively, with the major terpebiods being (+)-bornyl acetate, (+)-camphor, (+)-borneol, limonene, camphene, α-pinene, and β-pinene (Supplemental Figure S1, Supplemental Tables S1 and S2). To further determine the stereo structures of BPP-related terpenoids in both plants, we used chiral columns and found that only (+)-camphor and (+)-borneol were present in both plants (Supplemental Figure S2). Additionally, the stereo structure of bornyl acetate was assumed to be (+)-bornyl acetate based on the biosynthetic pathway. OPLS-DA analysis showed that (+)-bornyl acetate, (+)-camphor, limonene, camphene, and (+)-borneol were mainly enriched in seeds, while α-pinene and β-pinene were mainly enriched in flowers and other tissues (Supplemental Figure S3). The quantification of the seven major terpenoids revealed that BPP-related terpenoids were mainly enriched in seeds, while α-pinene and β-pinene were mainly produced in flowers and leaves (Figure 1C and Supplemental Table S3). Notably, BPP-related terpenoids were present in all tissues of *W. longiligularis* except for leaves, whereas in *W. villosa* they were found only in seeds, pericarp, and roots, indicating that BPP-related terpenoids were more widely distributed in *W. longiligularis* tissues (Figure 1C). Moreover, the percentage content of the seven terpenoids were similar in both plants and mainly consisted of (+)-bornyl acetate, but the contents of six terpenoids other than β-pinene in the *W. villosa* seeds were about five times higher than *W. longiligularis* (Supplemental Figures S4 and S5). These results demonstrate that the majority of terpenoids are shared between both plants and enriched

in seeds, but the major volatile terpenoid ((+)-bornyl acetate, (+)-camphor, (+)-borneol, limonene, and camphene) contents are significantly higher in *W. villosa* seeds.

Assembly and annotation of the *W. longiligularis* genome

To investigate the genetic basis of volatile terpenoids divergence between *W. longiligularis* and *W. villosa*, we performed BGI short-read, PacBio HiFi long-read, and Hi-C sequencing of *W. longiligularis* (Supplemental Table S4). Based on the K-mer analysis of the short-read sequencing data, we estimated the genome size of *W. longiligularis* to be 2.42 Gb, with a large proportion of repetitive sequences (78.8%) and relatively low heterozygosity (0.23%) (Supplemental Figure S6). A *de novo* assembly of the HiFi sequencing data was performed by Hifiasm, giving rise to an initial assembly of 2,293.25 Mb containing 320 contigs (contig N50 value: 80.39 Mb) (Supplemental Table S5). Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis using the embryophyta_odb10 database indicated that the genome completeness is 98.7% (Supplemental Table S6). We further used Hi-C data to anchor the assembled contigs to 24 pseudochromosomes with sizes ranging from 56.83 Mb to 125.52 Mb, which covered 98.76% (2,265.43 Mb) of the assembled genome size (Supplemental Table S7). The Hi-C contact map suggests a high quality of the chromosome-level assemblies (Supplemental Figure S7). In addition, 94.36% of the RNA-seq data could be mapped back to the genome, indicating a high accuracy of the assembled sequences (Supplemental Table S8). Finally, we obtained a chromosome-level genome of *W. longiligularis* containing 24 chromosomes with a total size of approximately 2.29 Gb (Figure 2A).

With a combination of *ab initio* gene prediction, homology evidence, and RNA-seq data,

we annotated 44,837 protein-coding genes in the genome of *W. longiligularis* with an average gene length of 5,017 bp and an average CDS length of 1,121 bp. BUSCO analysis indicated that the completeness of genome annotation is 93.8% using the embryophyta_odb10 database (Supplemental Table S6). A total of 43,718 (97.5%) protein-coding genes were functionally annotated through at least one of the following protein-related databases, including NCBI NR database (96.1%), eggNOG database (93.0%), InterPro domains (79.4%), KOG categories (29.9%), GO terms (64.7%), and KEGG pathways (25.2%). We also annotated 1,960 tRNA, 542 rRNA, 304 miRNA, and 5,506 snRNA in the *W. longiligularis* genome (Supplemental Table S9).

Comparative genomics analysis

To investigate the evolution history of the *W. longiligularis* genome, a comparative genomic analysis was conducted with four *Zingiberaceae* plants (*W. villosa*, *L. tsao-ko*, Ginger, and *C. alismatifolia*), two *Cannaceae* plants (*Canna edulis* and *Canna indica*), two *Musaceae* plants (banana (*Musa acuminate*) and *Musa schizocarpa*), a Arecales plant (areca palm (*Areca catechu*)), a Poales plant (rice (*Oryza sativa*)), and two outgroup plants (*Arabidopsis thaliana* and Grape (*Vitis vinifera*)). A total of 35,115 orthogroups (457,588 genes) were identified, including 386 single-copy genes shared by all 13 species (Supplemental Table S10). In addition, 7,578 orthogroups (18,674 genes) were shared by all five *Zingiberaceae* species, which likely represents the core genome of the ginger family (Supplemental Figure S8). Furthermore, the gene family evolutionary analysis revealed that 596 and 1,227 gene families underwent expansion and contraction, respectively, during the recent evolution of *W. longiligularis* after

its separation from *W. villosa* (Figure 2B). By contrast, a significantly higher number of gene families expanded in the *W. villosa* lineage. KEGG analysis of the expanded genes revealed substantially enrichment in terpenoid backbone biosynthesis, sesquiterpenoid and triterpenoid biosynthesis, indicating possible enhancements in the biosynthesis of volatile terpenoids (Supplemental Figure S9). These expanded genes may lead to differences in important traits and secondary metabolites between the two species.

To analyze the evolutionary relationship of the *Zingiberaceae*, a time-calibrated phylogenetic tree was constructed using the 386 single-copy genes, showing that the *Zingiberaceae* and *Cannaceae* diverged from their common ancestor ~66 million years ago (Mya) (Figure 2B). Notably, *Z. officinale* is sister to *C. alismatifolia* and followed by Wurfbainia plants (*L. tsao-ko*, *W. villosa*, and *W. longiligularis*), indicating that the Zingiber was more closely related to the Curcuma (Figure 2B). In addition, we inferred that the ancestors of *W. longiligularis* and *W. villosa* separated ~2.6 Mya, whereas the divergence between Wurfbainia and Zingiber occurred ~15 Mya (Figure 2B). The divergence times of *W. longiligularis* from *W. villosa* and *Z. officinale* were approximately predicted at 2.6 (*Ks*=0.06) and 15.0 (*Ks*=0.15) Mya, respectively, simultaneously, they produced the same signature *Ks* peaks of WGDs at about 0.35, higher than 0.06 and 0.15, indicating that the most recent WGD event occurred before the divergence of *Zingiberaceae* (Figure 2C). Intra-specific analyses of *W. longiligularis*, *W. villosa*, and *Z. officinale* all revealed the same peak of *Ks* distribution at 0.35, in between the peaks for interspecific comparison of *W. longiligularis* vs. *Z. officinale* (0.15) and *W. longiligularis* vs. *C. edulis* (0.69), indicating a WGD event in the ancestor of

Zingiberaceae (Figure 2C). Moreover, the branch lengths of the species tree suggest that *Zingiberaceae* evolved faster than *Musaceae* and *Cannaceae*, which further explains why the *Ks* value of *M. acuminata* vs. *C. edulis* was smaller than those of *W. longiligularis* vs. *M. acuminata* and *W. longiligularis* vs. *C. edulis* (Supplemental Figure S10).

To compare the genomic differences between *W. longiligularis* and *W. villosa*, 74,593 collinear genes were identified between the two plants, accounting for 85.3% of the total number of genes in two genomes. The inter-genomic analysis showed relatively conservative collinearity between *W. longiligularis* and *W. villosa*, indicative of a close evolutionary relationship (Supplemental Figure S11). Furthermore, we detected large structural variants (SVs) between the two genomes (Supplemental Table S11). Compared with the *W. villosa* genome, 234 regions were identified as inversions in *W. longiligularis* with a cumulative length of 540.9 Mb. Additionally, the *W. longiligularis* genome contained 2,218 translocations, spanning a total length of 62.8 Mb. Consequently, extensive interspecific genomic variations were identified, which could potentially lead to considerable variations in metabolites and phenotypic traits between *W. longiligularis* and *W. villosa*.

Comparative analysis of transposable elements

Although *W. longiligularis* and *W. villosa* diverged only ~2.6 Mya, their genome size differed by ~500 Mb. To understand the dynamics of genome size in Zingiberales, we annotated and compared the transposable elements (TEs) of six Zingiberales species (*W. longiligularis*, *W. villosa*, *Z. officinale*, *C. alismatifolia*, *C. edulis*, and *M. acuminata*) and discovered that: 1) *W. longiligularis* (1.93 Gb, 85.27%) and *W. villosa* (2.45 Gb, 87.14%) had the greatest TE

contents; and 2) the different genome sizes of *W. longiligularis* and *W. villosa* can be almost entirely attributed to TEs (Supplemental Table S12). In addition, the TE contents of *Zingiberaceae* species were all over 70% and significantly higher than those of *C. edulis* (*Cannaceae*) and *M. acuminate* (*Musaceae*), indicating a rapid expansion of TEs in *Zingiberaceae*. Insertion time analysis revealed that the expansion of the *Copia* and *Gypsy* superfamilies in the *W. longiligularis*, *W. villosa*, *Z. officinale*, and *C. alismatifolia* genomes exploded mainly within the last 2 Mya, indicating that these insertions occurred after the differentiation of these species (~15 Mya) (Figure 2D). In the *W. longiligularis* genome, the burst of LTR/*Copia* was stronger than LTR/*Gypsy*, indicating that the expansion of LTRs, particularly *Copia*, may contribute the most to the genome expansion. These results indicate that most of the LTRs in these six genomes were recent insertions and that the *Copia* and *Gypsy* burst were the major drivers of genome expansion.

Identification of *TPS* genes in *W. longiligularis*

TPSs are key enzymes for terpenoid biosynthesis and structural diversity and are responsible for catalyzing the generation of monoterpenoids, sesquiterpenoids, diterpenoids and triterpenoids. Strikingly, a total of 75 putative *WlTPSs* were identified in the *W. longiligularis* genome, which was similar to the *W. villosa TPS* genes (Supplemental Table S13). Phylogenetic analyses of TPS from *W. longiligularis*, *W. villosa*, and *O. sativa* indicated that the 75 WlTPSs were grouped into five previously recognized subfamilies, including TPS-a (40), TPS-b (23), TPS-c (3), TPS-e/f (5), and TPS-g (4) (Figure 3A). Notably, more than 80% of the WlTPSs belong to the TPS-a and TPS-b subfamilies, suggesting that substantial expansion of these two

subfamilies has a major contribution to the increased monoterpenoid and sesquiterpenoid biosynthesis. Chromosomal localization revealed that the 75 *WlTPSs* were unevenly distributed in the *W. longiligularis*; about two-thirds of the genes were located on three chromosomes, including Chr23 (22 *WlTPS*s), Chr22 (15 *WlTPS*s), and Chr14 (11 *WlTPS*s), whereas 11 of the 24 chromosomes do not carry any TPS gene (Supplemental Figure S12). In addition, 60 of the *WlTPS*s were located in a total of 14 tandem duplication clusters, the sizes of which range from 2 to 13 (Supplemental Figure S12). Collinearity analysis identified 34 *WlTPS* gene pairs located in syntenic blocks, indicating that segemental/whole-genome duplication process was also critical for the expansion of *W. longiligularis TPS* gene family (Supplemental Figure S12). These results suggest that *WlTPSs* have expanded through both large-scale and tandem duplication events.

Expression patterns of *WlTPSs* across tissues demonstrated that a total of 31 genes exhibited higher transcript abundance in seeds and flowers, which is consistent with the abundance of volatile terpenoids in *W. longiligularis* seeds and flowers (Figure 3B). In particular, *WlTPS1/30/73* was highly expressed in seeds (TPM>100), while *WlTPS26/28/51* showed even higher transcript abundance in flowers (TPM>1000), indicating that these genes may play a critical role in volatile terpenoid biosynthesis in *W. longiligularis* seeds and flowers (Supplemental Table S13). To further identify the key genes for terpenoid biosynthesis, we performed Pearson correlation analysis of co-expressed gene modules identified by WGCNA and the accumulation profiles of seven major terpenoid contents. We found 18 modules that were significantly correlated with at least one terpenoid. In particular, the "green" module was

correlated with the largest number of six terpenoids (Figure 3C). Notably, some important gene copies from the TPS gene family were grouped into the "green" module, including *WlTPS30*, *WlTPS43*, and *WlTPS73*, indicating that the expression patterns of these genes were similar to the accumulation trends of major terpenoids in seeds, and may be key genes responsible for the synthesis of volatile terpenoids in seeds (Figure 3D). In addition, functional annotation of the TPS gene subnetwork in the "green" module identified two MYB genes (*WL01Gene41708* and *WL01Gene43989*), suggesting that they may play an important role in the regulation of terpenoid biosynthesis (Figure 3D).

Functional characterization of WlBPPS

To characterize the bornyl diphosphate synthase (BPPS) responsible for the biosynthesis of BPP, we identified a tandem cluster including nine *WlTPSs* in chromosome 14 of *W. longiligularis* genome using collinearity analysis based on *WvBPPS* identified from *W. villosa* (Figure 4A). We termed this the 'BPPS cluster' and the genes BPPS-like genes. Then, we successfully cloned *WlTPS24/26/28/30/32*, while the low relative expression levels of *WlTPS25/27/29/31* and the high sequence identity (>88%) of the *WlBPPS* cluster prevented the cloning of these four genes. Functional identification indicated that WlTPS24/26/28/30 could catalyze GPP to produce BPP (transformed to borneol by dephosphorylation), camphene, and limonene (Figure 4A), and WlTPS32 catalyzed GPP to produce α-pinene, β-pinene, and limonene. Notably, WlTPS30 and WvBPPS similarly produce borneol (82%) as the main product, while the main product of WlTPS24 was also borneol (55%) but the product ratio was lower (Figure 4A). Meanwhile, the reaction kinetic parameters of WlTPS30 and WvBPPS with

GPP as substrate revealed that the K_m value of WvBPPS was slightly larger than those of WlTPS30, but the catalytic efficiencies (K_{cat}/K_m) of these two enzymes were close (Supplemental Figure S13). Moreover, our transcriptome and RT-qPCR data showed that *WlTPS30* was expressed predominantly in seeds, while *WlTPS24* was expressed in multiple tissues and *WlTPS26/28* were expressed predominantly in flowers (Supplemental Figure S14 and Supplemental Table S14). Therefore, we consider that WlTPS30 is the BPPS in *W. longiligularis* and named it WlBPPS; the gene is mainly responsible for the BPP biosynthesis in seeds, whereas WlTPS24/26/28 are involved in the BPP biosynthesis in other tissues.

GCN4-motif element drives high expression of *WvBPPS* in the seeds

Besides protein sequence divergence that may directly modulate BPPS activity, alterations in promoter sequences may lead to changes in gene expression status. Previously, we cloned the promoter sequences of *WlBPPS* and *WvBPPS*, respectively, and found that the *WvBPPS* promoter had a GCN4-motif (TGAGTCA) involved in endosperm-specific expression which didn't present in the promoter of *WlBPPS* but was replaced with 'C--GTCA' (Figure 4B) (Lin et al., 2022). In this study, we further investigated the cis-acting element in the promoter region of the BPPS-cluster genes in *W. longiligularis*. Interestingly, none of the BPPS-like genes except *WvBPPS* contained GCN4-motif; instead, the promoters of these genes mostly contain cis-acting elements associated with methyl jasmonate, abscisic acid, salicylic acid, light response, and low-temperature (Supplemental Figure S15). Thus, we speculate that the GCN4-motif element may affect the expression of *WlBPPS* and *WvBPPS* in the seeds.

To further investigate the function of GCN4-motif element, we generated recombinant

vectors containing different types of *WvBPPS* and *WlBPPS* promoters, and tested their activity in transgenic *N. tabacum* (Figure 4C). Using RT-qPCR analysis, we observed that the full-length *WvBPPS* promoter (VBP::GUS) predominantly drove *GUS* gene expression in seeds and leaves, while the full-length *WlBPPS* promoter (LBP::GUS) drove *GUS* gene expression mostly in leaves. Next, we truncated both promoters to their conserved regions (Figure 4B) and found that the truncated *WvBPPS* promoter (VBPT::GUS) showed high specificity in driving *GUS* gene expression in seeds, while the truncated *WlBPPS* promoter (LBPT::GUS) displayed low levels of *GUS* gene expression in seeds. Furthermore, we mutated the GCN4-motif element in VBPT::GUS and found that the *GUS* gene of VBPT-GM::GUS (*WvBPPS* truncated and mutated promoter) loss its specificity of expression in seeds. Notably, the expression of VBPT::GUS driven *GUS* gene was significantly higher than other genotypes, and the expression level of VBPT-GM::GUS driven *GUS* gene was significantly lower (Figure 4D). Additionally, GUS staining in various tissues of transgenic *N. tabacum* supported the RT-qPCR results (Supplemental Figure S16). These results suggest that the GCN4-motif element positively regulates the specific expression of *WvBPPS* in seeds, promoting the accumulation of BPP-related terpenoids in *W. villosa* seeds.

Functional characterization of WlTPSs

To further investigate the members of WlTPS involved in monoterpenoid biosynthesis, we cloned and functionally characterized eight additional TPS-b subfamily genes, which were expressed predominantly in seeds (Figure 5 and Supplemental Figure S17). WlTPS5 and WlTPS8 were able to catalyze the formation of α-pinene, β-pinene, limonene, and linalool from

GPP in different ratios, with the main product of WlTPS5 being α-pinene (60.5%) and that of WlTPS8 being β-pinene (68.2%) (Supplemental Table S15). WlTPS2 and WlTPS33 showed the ability to form α-ocimene and β-ocimene from GPP in varying proportions, with the primary product of WlTPS21 being α-ocimene (61.1%) and that of WlTPS33 being β-ocimene (66.2%) (Supplemental Table S15). Interestingly, WlTPS34/35/36/51 predominantly produced linalool when GPP was used as a substrate (Supplemental Table S15). In addition, only WlTPS36/51 of all TPS-b subfamily genes were able to catalyze the production of nerolidol from FPP (Supplemental Figure S18). These findings suggest that the functional diversity and distinct tissue expression patterns of TPS-b subfamily genes contributed to the high abundance and diverse array of monoterpenoids of *W. longiligularis*.

Next, WlTPS39/54/73 were classified under the TPS-a subfamily, and catalyzed FPP to form multiple sesquiterpenoids (Figure 5 and Supplemental Figure S18). Interestingly, when using GPP as a substrate, all three WlTPSs produced a variety of monoterpenoids, indicating that the majority of TPS-a possess bifunctionality in *W. longiligularis*. Furthermore, WlTPS1 belongs to the TPS-g subfamily and was capable of catalyzing FPP to exclusively produce nerolidol, as well as GPP to exclusively produce linalool (Supplemental Tables S15 and S16). Among these bifunctional WlTPSs, our transcriptome and RT-qPCR data indicated that *WlTPS73* was predominantly expressed in seeds, presumably playing a role in the biosynthesis of sesquiterpenoids in seeds (Supplemental Figure S14 and Supplemental Table S14).

Evolutionary analysis of *TPSs* and *BPPSs* in *Zingiberaceae*

Zingiberaceae plants have rich volatile terpenoids, mainly comprised of monoterpenoids

342 and sesquiterpenoids. To explore the evolution of the TPS gene family in *Zingiberaceae* plants
343 and to better understand terpenoid biosynthesis in monocot plants, we identified candidate TPSs
344 in 29 monocot plants from 16 families. *Zingiberaceae* species generally have more copies of
345 TPS genes than other monocots (Figure 6A). In particular, *W. longiligularis* and *Z. officinale*
346 encode the 75 and 74 *TPS* genes, respectively, representing the largest TPS repertoire in
347 monocots. In addition, the *TPS* genes were divided according to their subfamilies into 5 clades:
348 TPS-a, b, c, e/f, and g. Notably, our gene family analysis showed that TPS-a and TPS-b
349 subfamily genes are substantially expanded in the *Zingiberaceae* compared to all other monocot
350 lineages, indicating that these genes may promote a richer and more diverse set of volatile
351 terpenoids in the *Zingiberaceae* (Figure 6A).

352 Our sequence and functional analyses have demonstrated that BPPS plays an important
353 role in the biosynthesis of BPP-related terpenoids in *W. longiligularis* and *W. villosa*. To further
354 reveal the potential evolutionary history of *BPPS*, we performed a genomic syntenic analysis
355 of eight species of the Zingiberales (*Zingiberaceae*, *Cannaceae*, and *Musaceae*). Comparative
356 analysis of microsynteny blocks showed that *BPPS* were located in the genome collinearity
357 region of three different family plants, but not in the collinearity regions of species outside
358 Zingiberales, such as Arecales and Poales (Figure 6B). We termed the genes in all species that
359 are syntenic to *WvBPPS* as BPPS-like genes. Chromosomal localization revealed that these
360 BPPS-like genes are present as tandem duplication clusters in different species (Supplemental
361 Figure S19). Notably, we found no BPPS-like tandem duplication genes in the *BPPS* gene
362 collinearity blocks, suggesting that these tandem duplication genes may have been generated

363 by direct replication. This distribution pattern further suggests that *BPPS* appeared in
364 Zingiberales species before the divergence of *Zingiberaceae*, *Cannaceae*, and *Musaceae*,
365 indicating that BPPS genes were inherited from the common ancestor of Zingiberales species
366 and were highly conserved.

367 To test whether these BPPS-like genes have experienced functional divergence, enzyme
368 activities of recombinant LOC121970598 (*Z. officinale*), evm.model.LG16.2352 (*C.*
369 *alismatifolia*), MRJ010004766 (*C. indica*), BJY010017282 (*C. edulis*), Ms07t204890 (*M.*
370 *schizocarpa*), and Macma4_07_g28750 (*M. acuminata*) were performed *in vitro* (Figure 6C).
371 When GPP was used as a substrate, WvBPPS, WlTPS26, and evm.model.LG16.2352 in
372 *Zingiberaceae* were able to catalyze the biosynthesis of BPP, whereas LOC121970598 does not
373 have the catalytic activity. Notably, MRJ010004766 of the *Cannaceae*, but not BJY010017282,
374 catalyzed GPP to produce terpinene and terpinolene. Similarly, Macma4_07_g28750 of the
375 *Musaceae*, but not Ms07t204890, catalyzed GPP to generate α-ocimene. These results suggest
376 that the enzymatic activities of the BPPS-like genes were relatively conserved in the
377 *Zingiberaceae*, whereas functional differentiation has occurred in *Cannaceae* and the *Musaceae*.

378 **Site-directed mutagenesis of WvBPPS**

379 BPPS is the terpene cyclases in the TPS gene family responsible for catalyzing the
380 generation of BPP from GPP, and its activity directly affect the content of BPP-related
381 terpenoids in plants. In this study, comparison of the relative activities of WvBPPS, SoBPPS,
382 CbBPPS, LdBPPS, LaBPPS, and WlBPPS-like revealed that WvBPPS, WlBPPS, and SoBPPS
383 had the best activities, which could provide a reference for the heterologous synthesis of BPP-

384 related terpenoids (Figure 6D). To estimate key residues for functional differences in BPPS, we
385 integrated homology modeling and molecular docking analysis based on SoBPPS to identify
386 20 residues in the WvBPPS binding pocket (Supplemental Figure S20). Sequence alignment
387 showed that six residues in conserved motifs (RXR, DDXXD, and DTE/NSE) and seven
388 residues in lesser conserved motifs but conserved in all five BPPSs (Supplemental Figure S21).
389 Notably, a total of seven residues (I341, V345, V450, T451, V455, F488, and Y574) were not
390 fully conserved in all five BPPSs, suggesting that the interspecific differences in the activity of
391 BPPS are likely due to these variable residues in binding sites (Supplemental Figure S21).
392 Therefore, these seven residues were selected as targets of site-directed mutagenesis for
393 structural functional analysis.

394 Next, we verified the *in vitro* functions of seven incompletely conserved candidate residues
395 in WvBPPS by site-directed mutagenesis (Supplemental Figure S22). Compared with wild-type
396 WvBPPS, the mutation of I341R, V450A, and T451R showed 85.9%, 42.0% and 94.1% lower
397 activity against GPP, respectively, along with substantial changes in the ratios of products of
398 borneol, camphene and limonene (Figure 6E and Supplemental Table S17). In addition, the
399 activity of mutants V345R and F488W did not change considerably. Notably, the V455R and
400 V574R mutants of WvBPPS completely lost enzyme activity compared to the wild type,
401 suggesting that these two residues play an important role in substrate binding (Figure 6E and
402 Supplemental Table S17). Interestingly, we found that the 455th amino acid of the more active
403 BPPS was Val (V), while that of the less active BPPS was Ile (I) (Supplemental Figure S21).
404 To test the relationship between BPPS activity and the 455th amino acid, we mutated the Val of

WvBPPS to Ile and vice versa to Ile of WlTPS24 to Val. Functional characterization showed that the activity of the mutant V450I (WvBPPS) was reduced by 57.9%, while the activity of mutant I451V (WlTPS24) did not change considerably, indicating that mutation of a single key residue can substantially reduce the activity of BPPS while the increase in activity may be associated with other residues (Figure 6E and Supplemental Table S17). These results suggest that seven incompletely conserved key sites can provide a reference for studying the structure-function relationship of BPPS.

DISCUSSION

High-quality genomic resources of *W. longiligularis* have the potential to be utilized for varietal improvement of *Fructus Amomi*

W. longiligularis is a type of traditional Chinese medicine known for its diverse medicinal components, especially volatile terpenoids, which make it valuable for a wide range of applications in medicine and food (Doh et al., 2020). In this study, we produced a high-quality chromosome-level genome assembly of *W. longiligularis* by using PacBio HiFi long-read and Hi-C sequencing technologies. Compared to the published genomes of six other *Zingiberaceae* species, *W. longiligularis* had a significantly higher contig N50 value (80.39 Mb) than all others, such as *Z. officinale* (4.68 Mb), *W. villosa* (9.13 Mb), and *L. tsao-ko* (4.80 Mb), as well as *W. longiligularis* exhibited the highest BUSCO completeness (98.7%) (Li et al., 2021; Yang et al., 2022; Li et al., 2022). Repeated sequence analysis showed that the difference in genome size between *W. longiligularis* and *W. villosa* was primarily due to the proliferation of TEs, especially LTRs, rather than a polyploid event. Moreover, collinearity and SV studies revealed

the sequence similarity between the genes of *W. longiligularis* and *W. villosa*, despite large differences in their genomic structures. Based on these findings, we suggest that the discrepancy in plant phenotypes, environmental adaptations, and metabolite contents between *W. longiligularis* and *W. villosa* may be attributed to the variation in genome size caused by TEs amplification and large SVs.

Functional divergence of BPPS cluster genes and GCN4-motif element drive differences of BPP-related terpenoids in *W. longiligularis* and *W. villosa*

W. longiligularis and *W. villosa* are two distinct sources of *Fructus Amomi*, both of which are rich in volatile terpenoids, primarily bornyl acetate, camphor, and borneol (Ao et al., 2019). This study confirms that BPP-related terpenoids are more prevalent in the *W. villosa* seeds and have a wider tissue distribution in *W. longiligularis*. To investigate the genetic basis of the main monoterpeneoids divergence between *W. longiligularis* and *W. villosa*, we conducted a correlation analysis between different tissues, major TPSs, and seven main monoterpenoids of the two plants (Figure 7). Our findings show that in *W. longiligularis*, both WlBPPS and WlTPS24/26/28 can catalyze GPP to generate BPP, camphene, and limonene and are expressed in a variety of tissues, while in *W. villosa*, WvBPPS is the only TPS producing BPP and is expressed only in the seeds, which may explain the difference in BPP-related terpenoids of the two plants (Yang et al., 2022; Zhao et al., 2021). Additionally, we compared the functions of other TPSs in two plants and found that TPSs with similar evolutionary relationships have similar functions *in vitro* (Supplemental Table S18). For example, the pinene synthases in the two plants have similar functions *in vitro* and were expressed in all tissues, resulting in the

widespread distribution of α-pinene and β-pinene. In recent years, many studies have also shown that neofunctionalization or subfunctionalization of tandem duplication TPS genes in plants contributes to the diversity of terpenoids (He et al., 2022; Li et al., 2021; Wang et al., 2021). In this study, our findings suggest that the differential expression and functional divergence of BPPS cluster genes are responsible for the differences in the tissue distribution of BPP-related terpenoids between *W. longiligularis* and *W. villosa*. These results could serve as a reference for subsequent exploration of BPPS gene expression differences through epigenetic and transcriptional regulation.

Our findings demonstrate the substantial role of transcriptional regulation in determining the terpenoid composition and distribution of the two Wurfbainia plants. GCN4-motif is a conserved cis-element in the promoter of cereal crop seed storage proteins, controlling endosperm-specific expression of proteins like prolamin and glutelin (Wu et al., 2000; Kim et al., 2017). GCN4-motif acts in minimal-element combinations with ACGT-motif, AACA-motif, or the bZIP-like transcription factor *RISBZ1* to regulate glutenin-specific gene expression in the endosperm (Onodera et al., 2001). In contrast to the glutenin promoter, we showed that the GCN4-motif positively regulates the non-storage protein WvBPPS to achieve high expression levels in seeds in transgenic *N. tabacum*. Conversely, the promoter sequence of *WlBPPS* in *W. longiligularis* lacks the GCN4-motif element, resulting in low expression levels in seeds. Furthermore, we analyzed the promoter sequences of BPPS syntenic genes in *Zingiberaceae* plants and found that the GCN4-motif appears only in the promoter of *WvBPPS* (Supplemental Figure S23). This may be specific to the evolution of *W. villosa*, resulting in its seeds enriched

in BPP-related terpenoids. Therefore, our study provides a potential explanation for the higher content of BPP-related terpenoids in *W. villosa* seeds than *W. longiligularis*, which could provide genetic elements for improving the quality of *Fructus Amomi* through genetic engineering.

BPP-related terpenoids may be distributed only in the *Zingiberaceae* of monocot plants.

Borneol, camphor, and bornyl acetate are widely distributed in various plants (Timalsina et al., 2021; Mohammadhosseini et al., 2022; Cheng et al., 2022; Karpinski et al., 2020; Wang et al., 2020). However, the corresponding BPPS genes have been reported only in specific plant families, including *Zingiberaceae* (*WvBPPS*), *Lamiaceae* (*SoBPPS*, *SfBPPS*, *LaBPPS*, and *LiBPPS*), *Lauraceae* (*CbBPPS*), *Verbenaceae* (*LdBPPS*), and *Dipterocarpaceae* (*DzTPS3*, *DiTPS2*, *DgTPS1*, and *DaTPS3*) (Wang et al., 2018; Tian et al., 2022; Whittington et al., 2002; Czechowski et al., 2022; Ma et al., 2021). Notably, these three BPP-related terpenoids have been identified so far only in monocots from the *Zingiberaceae*, with WvBPPS being the sole BPPS gene identified in *W. villosa*. Our analysis suggests that the genomic neighborhood of *BPPS* gene was relatively conserved in Zingiberales of monocots, but not across other orders. Moreover, functional data indicate that the activities of BPPS-like genes are also relatively conserved in *Zingiberaceae*, with loss of activity or differentiation into TPSs with diverse functions observed in *Cannaceae* and *Musaceae*. These findings further support the idea that borneol and its related terpenoids may be unique to *Zingiberaceae* of monocots. Strangely, *Z. officinale* contains a minimal amount of borneol and camphor, but its BPPS putative ortholog (LOC121970598) apparently lacks activity *in vitro*, suggesting that the biosynthesis of BPP in

Z. officinale may be derived from other TPS by-products. In addition, the syntenic gene of *WvBPPS* in *W. longiligularis* is *WlTPS26* rather than *WlBPPS*, indicating that *WlBPPS* might have originated from *WlTPS26* by tandem duplication. Thus, the function and species-specific diversity of TPS may be more complicated than previously acknowledged, and the determination of the generated products based on sequence identity may be challenging.

Materials and Methods

Plant materials

The plant materials of *Wurfbainia longiligularis* were collected from the greenhouse of South China Botanical Garden, Chinese Academy of Sciences, China. The plant materials of *Wurfbainia villosa* were collected from the greenhouse of Guangzhou University of Chinese Medicine, China. The fresh young leaves were collected to extract high-quality genomic DNA for genome sequencing. Root (R), rhizome (RZ), stem (St), leaf (L), flower (F), pericarp (P) of 60-DAF (days after flowering), and seeds (S) of 60-DAF (days after flowering) were collected for RNA-seq sequencing and GC-MS analysis. All collected plant materials were washed with ultrapure water immediately, frozen in liquid nitrogen, and stored at -80℃.

DNA extraction and sequencing

High-quality genomic DNA was extracted from the young leaves of *W. longiligularis* using a modified CTAB method. For short-reads sequencing, the paired-end libraries were constructed with fragments of 350 bp and sequenced on the BGI MGISEQ-2000 platform. For PacBio sequencing, SMRTbell libraries were constructed according to PacBio's standard protocol (Pacific Biosciences, CA, USA) using 15 kb preparation solutions and sequenced on

the PacBio Sequel II platform with three cells. For the Hi-C experiment, about 3 g young leaves were prepared for cells fix by formaldehyde with 1% (w/v) formaldehyde solution in MS buffer, followed by chromatin extraction, chromatin digestion (DpnII), DNA ligation, purification and fragmentation, and the final constructed libraries were sequenced on the BGI MGISEQ-2000 platform.

Genome assembly and chromosome construction

To survey the *W. longiligularis* genome, GCE v1.0.0 (https://github.com/fanagislab/GCE) and GenomeScope2.0 v1.0.0 (Ranallo-Benavidez et al., 2020) were used to estimate genome size based on the k-mer frequency distribution. *De novo* assembly of the HiFi reads were performed using Hifiasm v0.15.4 (Cheng et al., 2021) with parameters '-l 3'. To obtain chromosome-level assemblies, Hi-C clean reads were first mapped to the contig assembly of *W. longiligularis* using Juicer v1.6 (Durand et al., 2016), and then the locations and directions of contigs were preliminarily determined using 3D-DNA v.201008 (Dudchenko et al., 2017). To prevent excessive interruption, the result of the first iteration of 3D-DNA was used as input for Juicerbox v2.13.07 (Durand et al., 2016) and to fix misjoins and clustered contigs. Finally, the corrected contigs were anchored into pseudochromosomes with 3D-DNA and visualized the Hi-C contact map. Genome assembly accuracy and completeness was evaluated by BUSCOs v5.2.2 (Seppey et al., 2019) (embryophyta_odb10) and transcriptome data.

Gene prediction and function annotation

Protein-coding genes of *W. longiligularis* genome were predicted using the GETA pipeline (https://github.com/chenlianfu/geta) based on homology annotation, RNA-sequencing

evidence, and *de novo* predictions. For the homology annotation, we searched the five genomic protein sequences: *W. villosa*, *Z. officinale*, *M. acuminate*, *O. sativa*, and *A. thaliana*. For RNA-sequencing evidence, we collected RNA-seq data from roots, rhizomes, stems, leaves, flowers, pericarp, and seeds. AUGUSTUS v3.3.3 (Stanke et al., 2005) was used for *de novo* gene prediction. Functional annotations of protein-coding genes were conducted based on NCBI NR (https://www.ncbi.nlm.nih.gov/), SwissProt (http://www.uniprot.org/), eggNOG v5.0 (http://eggnog-mapper.embl.de/), and InterPro v86.0 (https://www.ebi.ac.uk/interpro/) databases. miRNA, rRNA, and snRNA genes were identified using Infernal v1.1.2 (Nawrocki et al., 2013) to search the Rfam v14.5 database (Kalvari et al., 2021). Additionally, tRNA genes were identified using tRNAscan-SE v2.0.2 (Chan et al., 2019).

Comparative genomics analysis

Orthologous and paralogous groups were identified from *W. longiligularis* and 12 other plants using OrthoFinder v2.3.3 (Emms et al., 2019) with default parameters. Phylogenetic tree was constructed based on the protein sequences of single copy orthologous from 13 species using the maximum likelihood method in RAxML v8.2.12 (Stamatakis A., 2014) with 1000 bootstrap replicates and visualized by iTOL (https://itol.embl.de/). MCMCtree from PAML package v4.9 (Xu et al., 2013) was used to estimated species divergence times. Three calibration points were obtained from the TimeTree database (http://www.timetree.org/): *A. thaliana-V. vinifera* (109.0-123.5 Mya), *A. thaliana-O. sativa* (143.0-174.8 Mya), and *Z. officinale-W. villosa* (10.4-15.6 Mya). Expansion and contraction of gene families were detected using CAFE v5.0 (Mendes et al., 2020). Syntenic blocks were performed by JCVI v.0.84

552 (https://github.com/tanghaibao/jcvi). The *Ks* values of paralogous or orthologous gene pairs
553 were used to predict whole-genome duplication (WGD) event and calculated using wgd v1.1.1
554 (Zwaenepoel et al., 2019). To identify SVs, we used Mummer v4.0
555 (https://github.com/mummer4/mummer) to align the genomes of *W. longiligularis* and *W.*
556 *villosa*. Meanwhile, SyRI v1.2 (https://github.com/schneebergerlab/syri) was used to identify
557 genomic translocations and inversions between the genomes of *W. longiligularis* and *W. villosa*.
558 In addition, large SVs were identified using the same method of Sun *et al* (Sun et al., 2018).
559 Visualization of genomic features and syntenic blocks in the *W. longiligularis* genome using
560 Circos v0.69 (https://github.com/vigsterkr/circos).

561 **Transposable elements annotation**

562 TEs were identified in six Zingiberales species using a *de novo* methods with Extensive
563 *de-novo* TE Annotator (EDTA) v2.0.0 pipeline (Su et al., 2021). Subsequently, the unknown
564 LTR sequences were further classified into *Copia* and *Gypsy* superfamilies using DeepTE to
565 obtain the final non-redundant TE library for annotation of TEs. To refine this annotation, the
566 genomic TEs of these six species were then annotated again using EDTA with the settings '--
567 step anno --anno 1'. LTR insertion times were estimated using the adapted *T=K/2r* formula for
568 intacat LTR sequences following the approach described in LTR_retriever, and r = nucleotide
569 substitution rate (1.3×10^{-8} per site per year).

570 **Volatile terpenoids GC-MS detection and quantification**

571 Freeze-dried *W. longiligularis* and *W. villosa* powders (20 mg) were extracted with 1 mL
572 hexane in an ultrasonic cleaner for 30 min and then incubated at 40°C for 40 min. GC-MS

analysis was performed using an Agilent 7890B Gas Chromatograph with a 5977B inert Mass Selective Detector (Agilent, United States). Helium was used as a carrier gas (1 mL/min) and then separated on HP-5 (30 m × 0.25 mm, 0.25 μm film thickness, Agilent) and CycloSil-B (30 m × 0.25 mm, 0.25 μm film thickness, Agilent) columns. The HP-5 column was used for the detection of enzymatic reaction products and volatile terpenoids, while the CycloSil-B column was only used for the stereo structure identification of (+)-bornyl acetate, (+)-camphor, and (+)-borneol. Chromatograph conditions were as follows: injector, 220°C, initial oven temperature, 50°C, held for 5 min, increased by 4°C min^{-1} to 110°C, then increased by 2.5°C min^{-1} to 170°C, and held for 2 min. Volatile terpenoids were identified by NIST17/Wiley275 Mass Spectral Library and comparison of retention time and MS/MS spectra of authentic standards (Supplemental Figure S24). The relative quantification of volatile terpenoids was calculated using Agilent MassHunter software based on peak area ratio. Orthogonal partial least squares-discriminant analysis (OPLS-DA) regression analysis were performed using the SIMCA 14.1. The quantification of seven major terpenoids (α-pinene, camphene, β-pinene, limonene, (+)-camphor, (+)-borneol, and (+)-bornyl acetate) was performed by external standard method based on characteristic extracted ion chromatogram (EIC) (Supplemental Table S19). The volatile terpenoid analysis was performed with three biological replicates.

Gene expression and co-expression analysis

Total RNA was extracted from different tissues of *W. longiligularis* using TRNzol Universal Kit (Tiangen, China). RNA-seq libraries were generated using the TruSeq RNA Library Preparation Kit (Illumina, USA) and sequencing on the Illumina NovaSeq platform

with 150 bp paired-end reads. Clean reads were obtained by fastp v0.23.1 (Chen et al., 2018) and aligned to the *W. longiligularis* genome using HISAT2 v2.2.1 (Kim et al., 2019). The expression levels of genes were quantified with transcripts per million (TPM) using FeatureCounts. Weighted gene co-expression analysis was performed on all expressed genes (TPM > 0) using the R package WGCNA. Pearson correlations were calculated between module eigengenes (MEs) and the content of the seven major volatile terpenoids to identify significant modules. The networks were visualized by Cytoscape v.3.8.0 (https://cytoscape.org/).

Genes related to terpenoid biosynthesis

To identify the candidate TPS genes, conserved domains of the TPS gene family (PF01397 and PF03936) were used to search against protein sequences of *W. longiligularis* and 28 other monocot plants using hmmsearch. Candidate TPS genes were further manually checked with the Pfam database (http://pfam.xfam.org) to confirm the complete structural domain. Multiple sequences were aligned using MUSCLE v3.8.31 (https://www.ebi.ac.uk/Tools/msa/muscle/) and the maximum-likelihood tree was constructed using IQ-TREE v1.6.12 (Nguyen et al., 2015) with 1000 ultra-fast bootstrap replicates. Tree visualization was performed on EVOLVIEW (https://www.evolgenius.info/evolview/). Chromosome distribution was analyzed using TBtools v1.098769 (Chen et al., 2020).

Enzymatic activities of TPSs

The complete open reading frame (ORF) of the *WlTPSs* were amplified with precise primers and then subcloned into the pLB cloning vector (Tiangen, China) (Supplemental Table S20). The prokaryotic expression vectors were constructed by ligating *WlTPS* ORFs into the

pET32a expression vector with the help of In-Fusion Cloning Kit (Takara, Japan) (Supplemental Table S21). We also synthesized *SoBPPS* (AF051900), *CbBPPS* (QTW43990), *LdBPPS* (ATY48638), *LaBPPS* (AJW68082), *LOC121970598*, *evm.model.LG16.2352*, *MRJ010004766*, *BJY010017282*, *Macma4_07_g28750*, and *Ms07t204890*, and constructed pET32a expression vectors through TsingkeBiotechnology Co., Ltd (Guangzhou, China). Positive constructs were transformed into competent Rosetta (DE3) cells and incubated at 37°C until the OD 600 reached 0.4-0.6. The recombinant proteins were induced using 0.1 mM IPTG at 16 °C overnight and purified with the NI-NTA resin (Qiagen, Hilden, Germany). Enzyme assays were carried out in 100 μL reaction mixture (25 mM HEPES, pH 7.2, 5 mM $MgCl_2$, and 5 mM DTT), which contained 20 μg purified protein and 50 μM substrate (GPP or FPP). The mixture was then incubated for 2 h at 30 °C and extracted with 200 μL of hexane. Analysis was conducted using the same method as for volatile terpenoid measurements. Samples were treated with 2.0 μL alkaline phosphatase (Thermo Fisher, United States) at 37 °C for an additional 1 h after the reaction to dephosphorylate BPP. Each enzyme assay was performed with at least three replicates.

To determine kinetic parameters, enzyme assays were carried out using 10 ng of purified protein and varying concentrations of GPP (1-150 μM). The reaction mixture was incubated at 30 °C for 10 minutes, followed by termination of the reaction using a water bath at 80 °C for 10 minutes. The other methods were the same as for the enzymatic reaction. The experimental procedures were repeated thrice. The reaction rate of the enzyme was determined by measuring the amount of the dephosphorylation product borneol. The kinetic parameters were then

636 determined by performing a nonlinear regression analysis by using GraphPad Prism 9 software.

637 **Construction of vector and plant transformation**

638 Six recombinant vectors were cloned into *BamHI* and *NcoI* sites of pCAMBIA1301 vector
639 and containing different promoter sequences of *WlBPPS* and *WvBPPS* that were pre-cloned in
640 our lab (Supplemental Table S22). The pCAMBIA1301 recombinant vector was created using
641 the In-fusion enzyme, fused with *GUS* gene, and then transferred into *E. coli* DH5α. Aseptic *N.*
642 *tabacum* leaves were pre-cultured and subsequently placed in a re-suspended Agrobacterium
643 tumefaciens solution (100 mM acetosyringone, 10 mM ethane sulfonic acid, and 10 mM
644 $MgCl_2$). The leaves were cultured in the dark on the co-culture medium (6-BA 1 mg/L and
645 pH=5.8) for 2 days. After co-culture, the resistant buds were cut and transferred into the rooting
646 medium (1/2MS, Timentin 300 mg/L, IAA 0.5 mg/L, HYG 15 mg/L and pH=5.8) when they
647 grew to 2 cm high. Upon forming complete plants, they were transferred to the soil.

648 **Molecular docking and site-directed mutagenesis**

649 SwissModel was used to perform homology modeling of WvBPPS based on the crystal
650 structure of *Salvia officinalis* SoBPPS (PDB ID: 1N24.1.A) (Whittington et al., 2002). The
651 reliability of the protein models was evaluated by Procheck and energy was minimized using
652 SPDBV (https://spdbv.unil.ch/). Molecular docking was conducted using AutoDock vina
653 (Eberhardt et al., 2021), using grid points with 50 × 50 × 50 spacing set at 0.375 Å. The resulting
654 complexes were visualized using PyMOL (https://pymol.org/2/). *WvBPPS* was mutated using
655 a site-directed mutagenesis kit (TIANGEN, Beijing, China) according to the site-directed
656 mutagenesis primers, which was then amplified using ProFlex PCR system (Supplemental

657 Table S23). PCR products were treated with *DpnI* restriction enzyme at 37 °C for 1 hour to
658 remove the template plasmid, and then transferred into competent Rosetta (DE3) cells. Finally,
659 recombinant proteins were purified and functionally identified using the method described
660 above.

661 **Reverse Transcription quantitative PCR**

662 Reverse transcription quantitative PCR (RT-qPCR) was used to performed the transcript
663 expression of *WlTPS* and different transgenic *N. tabacum GUS*. The primers used for RT-qPCR
664 were designed by Primer 5 based on the specific sequences of target gene (Supplemental Table
665 S24). The RT-qPCR reaction was conducted using 2×M5 HiPer SYBR Premix EsTaq (Mei5
666 Biotechnology, China) on a CFX 96 Real-Time PCR Detection System (BioRad, United States)
667 following the manufacturers' instructions. The thermal cycling conditions were 95°C for 30 s,
668 followed by 35 cycles of 95°C for 10 s and 58°C for 30 s. A single internal reference α-tubulin
669 gene was used as a standardized control to correct the transcription level of the target gene and
670 calculated using the $2^{-\triangle\triangle Ct}$ method. All experiments were performed with three biological
671 replicates.

672 **Accession Numbers**

673 The raw genome and transcriptome sequencing data reported in the present study have been
674 deposited in the National Center for Biotechnology Information (NCBI) database under project
675 number PRJNA813026. Additionally, the genome assembly, gene structure annotations,
676 predicted CDS, protein sequences, and WlTPSs protein and CDS sequences are available at
677 FigShare (https://doi.org/10.6084/m9.figshare.21579504).

Supplemental Data

Supplemental Figure S1. Extracted ion chromatogram (EIC) of monoterpenoids and sesquiterpenoids in *W. longiligularis* and *W. villosa*.

Supplemental Figure S2. Total ion chromatogram (TIC) for GC-MS analysis of *W. longiligularis* and *W. villosa* using the CycloSil-B column.

Supplemental Figure S3. OPLS-DA analysis of volatile terpenoids in *W. longiligularis* and *W. villosa*.

Supplemental Figure S4. Content distribution of seven terpenoids in *W. longiligularis* and *W. villosa* seeds.

Supplemental Figure S5. Comparison of the contents of seven terpenoids in *W. longiligularis* and *W. villosa* seeds.

Supplemental Figure S6. 21-mer analysis to estimate the *W. longiligularis* genome size.

Supplemental Figure S7. Heatmap of Hi-C interaction density between 24 pseudochromosomes of *W. longiligularis*.

Supplemental Figure S8. The shared and unique gene families were compared among five *Zingiberaceae* species.

Supplemental Figure S9. The KEGG pathway analysis of expanded genes in *W. longiligularis* genome.

Supplemental Figure S10. Phylogenetic tree of 13 plant species.

Supplemental Figure S11. Syntenic relationship between *W. longiligularis* and *W. villosa* chromosomes.

Supplemental Figure S12. Chromosome distribution and segmental duplication analysis of *WlTPSs*.

Supplemental Figure S13. Enzyme kinetic analysis of WvBPPS and WlBPPS with GPP as substrate.

Supplemental Figure S14. The expressional level of *WlTPS* in different tissues.

Supplemental Figure S15. Various cis-elements analysis of *WvBPPS* and WlBPPS-cluster genes.

Supplemental Figure S16. GUS staining of various tissues in transgenic *N. tabacum*.

Supplemental Figure S17. The GC-MS chromatograms of products generated by recombinant WlTPSs with added GPP substrate.

Supplemental Figure S18. The GC-MS chromatograms of products generated by recombinant WlTPSs with added FPP substrate.

Supplemental Figure S19. Chromosome distribution analysis of BPPS-like genes in Zingiberales species.

Supplemental Figure S20. Homology modeling and molecular docking analysis of WvBPPS.

Supplemental Figure S21. Multiple sequence alignment of WvBPPS, WlBPPS, SoBPPS, LaBPPS, CbBPPS, and WlTPS24.

Supplemental Figure S22. Functional characterization of various mutants of WvBPPS.

Supplemental Figure S23. Various cis-elements analysis of BPPS-like genes in Zingiberaceae.

Supplemental Figure S24. Mirror MS/MS spectra of the products (red) and standards (blue).

Supplemental Table S1. The volatile terpenoids in different tissues of *W. longiligularis*.

720 **Supplemental Table S2.** The volatile terpenoids in different tissues of *W. villosa*.

721 **Supplemental Table S3.** Statistics on the content (ng/mg) of 7 monoterpenoids in different
722 tissues of *W. longiligularis* and *W. villosa*. Errors represent mean ± SD (n = 3 biologically
723 independent samples).

724 **Supplemental Table S4.** Summary of sequencing data for *W. longiligularis*.

725 **Supplemental Table S5.** Overview of the genome assembly of *W. longiligularis*.

726 **Supplemental Table S6.** Evaluation of completeness of the final genome assembly and
727 annotation using BUSCO.

728 **Supplemental Table S7.** Summary of *W. longiligularis* chromosome-level assembly.

729 **Supplemental Table S8.** Percentages of RNA-seq reads mapped to the *W. longiligularis*
730 genome.

731 **Supplemental Table S9.** Statistics of non-protein-coding RNA annotations in the *W.*
732 *longiligularis* genome.

733 **Supplemental Table S10.** Statistics of gene families in 13 species.

734 **Supplemental Table S11.** Statistics of structural variations between *W. villosa* and *W.*
735 *longiligularis*.

736 **Supplemental Table S12.** Statistics of transposable elements in the genomes of *W.*
737 *longiligularis*, *W. villosa*, *Z. officinale*, *C. alismatifolia*, *C. edulis*, and *M. acuminata*.

738 **Supplemental Table S13.** Information on *WlTPSs* and their expression levels (transcripts per
739 million, TPM) in different tissues of *W. longiligularis*.

740 **Supplemental Table S14.** Data of the relative expression level (RT-qPCR) of *WlTPS* in

741 different tissues.

742 **Supplemental Table S15.** Percentages of monoterpenoid products with GPP as substrate.

743 **Supplemental Table S16.** Percentages of sesquiterpenoid products with FPP as substrate.

744 **Supplemental Table S17.** Percentages of WvBPPS mutant products with GPP as substrate.

745 **Supplemental Table S18.** Functional comparison of TPSs in *W. longiligulairs* and *W. villosa*.

746 **Supplemental Table S19.** Seven monoterpenoid standard curves.

747 **Supplemental Table S20.** Primers used for gene cloning.

748 **Supplemental Table S21.** Primers used for expression vector construction.

749 **Supplemental Table S22.** Primers used for site-directed mutation.

750 **Supplemental Table S23.** Primers used for RT-qPCR.

751 **Funding**

752 This work is financially supported by the National Natural Science Foundation of China
753 (81303163 and 81872954), Key-Area Research and Development Program of Guangdong
754 Province (No.2020B20221001 and No.2020B0202090001), and Hunan Provincial Key
755 Laboratory for Synthetic Biology of Traditional Chinese Medicine (No: HCSW2021-01).

756 Conflict of interest statement. None declared.

757

758 **Acknowledgments**

759 We thank Zhang Xudong and Wang Ying for their valuable help in genome analysis.

760

761 **Author contributions**

762 J.F.Y., P.Y., X.S.H., and R.T.Z. conceived and initiated the study; P.Y., X.F.Z., and Y.X.C.

performed the bioinformatics analyses; X.Y.L., T.T.W., X.J.L., and Y.Y.Z. performed most of the experiments; L.X.H, Y.W.S., and Y.X.Q. assisted in part of the experiment; Y.S.Y. collected plant material; P.Y. wrote the manuscript; J.F.Y., X.F.Z., and D.M.M. revised the manuscript.

Figure Legends

Figure 1 Distribution of volatile terpenoids in *W. longiligularis* and *W. villosa*. (A) Morphological characteristics of *W. longiligularis*. (B) Morphological characteristics of *W. villosa*. a: Plant; b: Flower; c: Fruit. (C) Content distribution of (+)-bornyl acetate, (+)-camphor, (+)-borneol, limonene, camphene, α-pinene, and β-pinene in different tissues of *W. longiligularis* and *W. villosa*.

Figure 2 Genomic features of *W. longiligularis*. (A) Circos plot of *W. longiligularis* genome assembly. The window is 1 Mb in size. a, Chromosome karyotypes; b, GC content (min-max, 0-1.0); c, Non-coding RNA (ncRNA) density (min-max, 0-0.002), d, Gene density (min-max, 0-80); e, Transposable elements density (<=0.6, green; 0.6~0.8, yellow; >=0.8, red); f, Long terminal repeat (LTR) density (<=0.6, green; 0.6~0.8, yellow; >=0.8, red); g, Syntenic blocks. (B) Phylogenetic tree and gene family expansions or contractions in 13 species. (C) Distribution of Ks values between *W. longiligularis*, *W. villosa*, *Z. officinale*, *M. acuminata*, and *C. edulis*. (D) Distribution of LTR (*Gypsy* and *Copia*) insertion time of *W. longiligularis*, *W. villosa*, *Z. officinale*, and *C. alismatifolia*.

Figure 3 Analysis of TPS gene family in *W. longiligularis*. (A) Phylogenetic tree of TPS genes from *W. longiligularis* (75 genes), *W. villlosa* (66 genes), and *O. sativa* (32 genes). Phylogenetic tree ignores branch length information. Red star means *W. longiligularis*, blue star means *W.*

villosa, and pink triangle means *O. sativa*. (B) Heatmap (row scale) showing the differential expression of *WlTPSs* according to the transcriptome data from various tissues (R, root; RZ, rhizome; St, stem; L, leaf; F, flower; P, pericarp; S, seeds). (C) Correlation analysis of WGCNA co-expression module with seven terpenoid contents in *W. longiligularis*. (D) Network analysis of TPS-related genes in the green module. The orange hexagon means the TPS genes in the green module.

Figure 4. Genetic basis of volatile terpenoids divergence in *W. longiligularis* and *W. villosa*. (A) Functional characterization and comparison of WlBPPS cluster genes in *W. longiligularis.* (B) Multiple sequence alignment of the *WvBPPS* and *WlBPPS* promoters. The blue boxes are the conserved regions of the promoter sequences of the two genes. (C) Analysis of *GUS* gene expression in different transgenic *N. tabacum*. The transgenic *N. tabacum* seeds were statistically analyzed with other tissues (unpaired t-test, *, $P < 0.05$; **, $P < 0.01$). Error bars represent mean ± SD (n = 3 biologically independent samples). (D) Comparison of *GUS* expression in seeds of different transgenic *N. tabacum*. VBPT::GUS compared to other transgenic types (unpaired t-test, *, $P < 0.05$; **, $P < 0.01$). Error bars represent mean ± SD (n = 3 biologically independent samples). VBP::GUS indicates the full-length promoter of *WvBPPS*, VBPT::GUS indicates the truncated to conserved region promoter of *WvBPPS*, VBPT-GM::GUS indicates the truncated to conserved region and GCN4-motif mutant promoter of *WvBPPS*, LBP::GUS indicates the full-length promoter of *WlBPPS* , and LBPT::GUS indicates the truncated to conserved region promoter of *WlBPPS*.

Figure 5 Major products of functional characterization *in vitro* of 17 WlTPSs in *W. longiligularis*. Rectangular background colors depict different subfamilies of WlTPSs: TPS-a (orange), TPS-b (blue), and TPS-g (gray). GPP, geranyl diphosphate; FPP, farnesyl diphosphate; 1, α-pinene; 2, β-pinene; 3, (+)-bornyl diphosphate; 4, camphene; 5, limonene; 6, myrcene; 7, linalool; 8, α-ocimene; 9, β-ocimene; 10, α-terpineol; 11, nerolidol; 12, 4a,8-dimethyl-2-(prop-1-en-2-yl)-1,2,3,4,4a,5,6,7-octahydronaphthalene; 13, β-selinene; 14, α-santalene.

Figure 6 Evolutionary dynamics of BPPS genes and site-directed mutagenesis of WvBPPS. (A) Expansion of TPS genes in *Zingiberaceae*. A total of 916 TPS genes were indentified in 29 monocot plants. Phylogenetic tree ignores branch length information. (B) Synteny analysis showing the conservation of BPPS in Zingiberales species. The blue line shows the syntenic relationship of BPPS among different Zingiberales species. (C) Functional characterization of syntenic genes in Zingiberales species. 1: α-pinene, 2: camphene, 3: terpinolene, 4: limonene, 5: terpinene, 6: α-ocimene, 7: (+)-borneol, 8: geraniol. (D) Relative activity of borneol produced by different plants BPPS and WlBPPS-like. Error bars represent mean ± SD (n = 3 independent experiments). ***$P < 0.001$; ns, no significant difference; unpaired t-test (n = 3). (E) Relative activity of borneol produced by various mutants. Error bars represent mean ± SD (n = 3 independent experiments). ***$P < 0.001$; ns, no significant difference; nd, not detected; unpaired t-test (n = 3).

Figure 7 Association analysis between different tissues, main TPSs, and seven main monoterpenoids of the *W. longiligularis* and *W. villosa*. Different color junctions represent TPS genes specific expression tissues.

References

Mei Y, Li L, Fan L, Fan W, Liu L, Zhang F, Hu Z, Wang K, Yang L, Wang Z (2023) The history, stereochemistry, ethnopharmacology and quality assessment of borneol. J Ethnopharmacol **300**: 115697

Zhao ZJ, Sun YL, Ruan XF (2023) Bornyl acetate: A promising agent in phytomedicine for inflammation and immune modulation. Phytomedicine **114**: 154781

Bahramikia S, Drikvandi P (2022) The abundance of camphor, as a plant essential oil, in various plants and their different parts. Nat Prod Res **3**: 1-2

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA **99**: 15375-80

Adal AM, Najafianashrafi E, Sarker LS, Mahmoud SS (2023) Cloning, functional characterization and evaluating potential in metabolic engineering for lavender (+)-bornyl diphosphate synthase. Plant Mol Biol **111**: 117-130

Hurd MC, Kwon M, Ro DK (2017) Functional identification of a *Lippia dulcis* bornyl diphosphate synthase that contains a duplicated, inhibitory arginine-rich motif. Biochem Biophys Res Commun **490**: 963-968

Wang H, Ma D, Yang J, Deng K, Li M, Ji X, Zhong L, Zhao H (2018) An integrative volatile terpenoid profiling and transcriptomics analysis for gene mining and functional characterization of AvBPPS and AvPS involved in the monoterpenoid biosynthesis in *Amomum villosum*. Front Plant Sci **9**: 846

Yang P, Zhao HY, Wei JS, Zhao YY, Lin XJ, Su J, Li FP, Li M, Ma DM, Tan XK, et al. (2022) Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of *Wurfbainia villosa*. Plant J **112**: 630-645

Doh EJ, Lee G, Jung HJ, Kwon KB, Kim JH (2020) Chemotaxonomic monitoring of genetically authenticated *Amomi Fructus* using high-performance liquid chromatography-diode array detector with chemometric analysis. Molecules **25**: 4581

Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C, et al. (2021) Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in *Salvia miltiorrhiza*. Nat Commun **12**: 685

Zhong C, Chen C, Gao X, Tan C, Bai H, Ning K (2022) Multi-omics profiling reveals comprehensive microbe-plant-metabolite regulation patterns for medicinal plant Glycyrrhiza uralensis Fisch. Plant Biotechnol J **20**: 1874-1887

Jia Q, Brown R, Köllner TG, Fu J, Chen X, Wong GK, Gershenzon J, Peters RJ, Chen F (2022) Origin and early evolution of the plant terpene synthase family. Proc Natl Acad Sci USA **119**: e2100361119

Bao T, Kimani S, Li Y, Li H, Yang S, Zhang J, Wang Q, Wang Z, Ning G, Wang L, et al. (2023) Allelic Variation of Terpene Synthases Drives Terpene Diversity in the Wild Species of the Freesia Genus. Plant Physiol **18**: kiad172

Rosenkranz M, Chen Y, Zhu P, Vlot AC (2021) Volatile terpenes-mediators of plant-to-plant communication. Plant J **108**: 617-631

Nagegowda DA, Gupta P (2020) Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. Plant Sci **294**: 110457

Guo X, Fang D, Sahu SK, Yang S, Guang X, Folk R, Smith SA, Chanderbali AS, Chen S, Liu M, et al. (2021) Chloranthus genome provides insights into the early diversification of angiosperms. Nat Commun **12**: 6930

Barbosa GB, Jayasinghe NS, Natera SHA, Inutan ED, Peteros NP, Roessner U (2017) From common to rare *Zingiberaceae* plants-A metabolomics study using GC-MS. Phytochemistry **140**: 141-150

Tunnisa F, Nur Faridah D, Afriyanti A, Rosalina D, Ana Syabana M, Darmawan N, Dewi Yuliana N (2022) Antioxidant and antidiabetic compounds identification in several indonesian underutilized *Zingiberaceae* spices using SPME-GC/MS-based volatilomics and *in silico* methods. Food Chem X **14**: 100285

Kulyal P, Acharya S, Ankari AB, Kokkiripati PK, Tetali SD, Raghavendra AS (2021) Variable secondary metabolite profiles across cultivars of *Curcuma longa* L. and *C. aromatica* Salisb. Front Pharmacol **12**: 659546

Peng W, Li P, Ling R, Wang Z, Feng X, Liu J, Yang Q, Yan J (2022) Diversity of volatile compounds in ten varieties of *Zingiberaceae*. Molecules **27**: 565

Ivanović M, Makoter K, Razboršek M (2021) Comparative study of chemical composition and antioxidant activity of essential oils and crude extracts of four characteristic *Zingiberaceae* herbs. Plants (Basel) **10**: 501

Li HL, Wu L, Dong Z, Jiang Y, Jiang S, Xing H, Li Q, Liu G, Tian S, Wu Z, et al. (2021) Haplotype-resolved genome of diploid ginger (*Zingiber officinale*) and its unique gingerol biosynthetic pathway. Hortic Res **8**: 189

Liao X, Ye Y, Zhang X, Peng D, Hou M, Fu G, Tan J, Zhao J, Jiang R, Xu Y, et al. (2022) The genomic and bulked segregant analysis of *Curcuma alismatifolia* revealed its diverse bract pigmentation. Abiotech **3**: 178-196

Yin Y, Xie X, Zhou L, Yin X, Guo S, Zhou X, Li Q, Shi X, Peng C, Gao J (2022) A chromosome-scale genome assembly of turmeric provides insights into curcumin biosynthesis and tuber formation mechanism. Front Plant Sci **13**: 1003835

Li P, Bai G, He J, Liu B, Long J, Morcol T, Peng W, Quan F, Luan X, Wang Z, et al. (2022) Chromosome-level genome assembly of *Amomum tsao-ko* provides insights into the biosynthesis of flavor compounds. Hortic Res **9**: uhac211

Lin XJ, Liang HL, Zhao HY, Wu QW, Huang LX, Wu SR, Yang JF (2022) Comparison of BPPS promoters between *Amomum villosum* and *Amomum longiligulare* and identification of GCN4 motif positive regulation. Chinese Traditional and Herbal Drugs **53**: 6159-6166

Ao H, Wang J, Chen L, Li S, Dai C (2019) Comparison of Volatile Oil between the Fruits of *Amomum villosum* Lour. and *Amomum villosum* Lour. var. *xanthioides* T. L. Wu et Senjen Based on GC-MS and Chemometric Techniques. Molecules **24**: 1663

Zhao H, Li M, Zhao Y, Lin X, Liang H, Wei J, Wei W, Ma D, Zhou Z, Yang J (2021) A Comparison of Two Monoterpenoid Synthases Reveals Molecular Mechanisms Associated with the Difference of Bioactive Monoterpenoids Between *Amomum*

villosum and *Amomum longiligulare*. Front Plant Sci **12:** 695551

He J, Verstappen F, Jiao A, Dicke M, Bouwmeester HJ, Kappers IF (2022) Terpene synthases in cucumber (*Cucumis sativus*) and their contribution to herbivore-induced volatile terpenoid emission. New Phytol **233**: 862-877

Li J, Wang Y, Dong Y, Zhang W, Wang D, Bai H, Li K, Li H, Shi L (2021) The chromosome-based lavender genome provides new insights into *Lamiaceae* evolution and terpenoid biosynthesis. Hortic Res **8**: 53

Wang X, Gao Y, Wu X, Wen X, Li D, Zhou H, Li Z, Liu B, Wei J, Chen F, et al. (2021) High-quality evergreen azalea genome reveals tandem duplication-facilitated low-altitude adaptability and floral scent evolution. Plant Biotechnol J **19**: 2544-2560

Wu C, Washida H, Onodera Y, Harada K, Takaiwa F (2000) Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal cis-element requirements for endosperm-specific gene expression. Plant J **23**: 415-21

Kim JS, Chae S, Jun KM, Pahk YM, Lee TH, Chung PJ, Kim YK, Nahm BH (2017) Genome-wide identification of grain filling genes regulated by the OsSMF1 transcription factor in rice. Rice **10**: 16

Onodera Y, Suzuki A, Wu CY, Washida H, Takaiwa F (2001) A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J Biol Chem **276**: 14139-52

Timalsina D, Devkota HP (2021) *Eclipta prostrata* (L.) L. (*Asteraceae*): ethnomedicinal uses, chemical constituents, and biological activities. Biomolecules **11**: 1738

Mohammadhosseini M, Frezza C, Venditti A, Mahdavi B (2022) An overview of the genus *Aloysia* Paláu (*Verbenaceae*): Essential oil composition, ethnobotany and biological activities. Nat Prod Res **36**: 5091-5107

Cheng KK, Nadri MH, Othman NZ, Rashid SNAA, Lim YC, Leong HY (2022) Phytochemistry, Bioactivities and Traditional Uses of *Michelia* × *alba*. Molecules **27**: 3450

Karpiński TM (2020) Essential Oils of *Lamiaceae* Family Plants as Antifungals. Biomolecules **10**: 103

Wang J, Su B, Jiang H, Cui N, Yu Z, Yang Y, Sun Y (2020) Traditional uses, phytochemistry and pharmacological activities of the genus Cinnamomum (Lauraceae): A review. Fitoterapia **146:** 104675

Tian Z, Zeng P, Lu X, Zhou T, Han Y, Peng Y, Xiao Y, Zhou B, Liu X, Zhang Y, et al. (2022) Thirteen *Dipterocarpoideae* genomes provide insights into their evolution and borneol biosynthesis. Plant Commun **3**: 100464

Czechowski T, Branigan C, Rae A, Rathbone D, Larson TR, Harvey D, Catania TM, Zhang D, Li Y, Salmon M, et al. (2022) *Artemisia annua* L. plants lacking Bornyl diPhosphate Synthase reallocate carbon from monoterpenes to sesquiterpenes except artemisinin. Front Plant Sci **13**: 1000819

Ma R, Su P, Guo J, Jin B, Ma Q, Zhang H, Chen L, Mao L, Tian M, Lai C, et al. (2021) Bornyl Diphosphate Synthase from *Cinnamomum burmanni* and its application for (+)-borneol biosynthesis in yeast. Front Bioeng Biotechnol **9**: 631863

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for

953 reference-free profiling of polyploid genomes. Nat Commun **11**: 1432
954 **Cheng H, Concepcion GT, Feng X, Zhang H, Li H** (2021) Haplotype-resolved de novo
955 assembly using phased assembly graphs with hifiasm. Nat Methods **18**: 170-175
956 **Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL** (2016)
957 Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell
958 Syst **3**: 95-8
959 **Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS,**
960 **Machol I, Lander ES, Aiden AP et al.** (2017) *De novo* assembly of the *Aedes*
961 *aegypti* genome using Hi-C yields chromosome-length scaffolds. Science **356**: 92-95
962 **Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL** (2016)
963 Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell
964 Syst **3**: 99-101
965 **Seppey M, Manni M, Zdobnov EM** (2019) BUSCO: assessing genome assembly and
966 annotation completeness. Methods Mol Biol **1962**: 227-245
967 **Stanke M, Morgenstern B** (2005) AUGUSTUS: a web server for gene prediction in
968 eukaryotes that allows user-defined constraints. Nucleic Acids Res **33**: W465-7.
969 **Nawrocki EP, Eddy SR** (2013) Infernal 1.1: 100-fold faster RNA homology searches.
970 Bioinformatics **29**: 2933-5
971 **Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M,**
972 **Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, et al.** (2021) Rfam 14:
973 expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res **49**:
974 D192-D200
975 **Chan PP, Lowe TM** (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences.
976 Methods Mol Biol **1962**: 1-14
977 **Emms DM, Kelly S** (2019) OrthoFinder: phylogenetic orthology inference for comparative
978 genomics. Genome Biol **20**: 238
979 **Stamatakis A** (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
980 large phylogenies. Bioinformatics **30**: 1312-3
981 **Xu B, Yang Z** (2013) PAMLX: a graphical user interface for PAML. Mol Biol Evol **30**: 2723-
982 4
983 **Mendes FK, Vanderpool D, Fulton B, Hahn MW** (2020) CAFE 5 models variation in
984 evolutionary rates among gene families. Bioinformatics btaa1022
985 **Zwaenepoel A, Van Y** (2019) wgd-simple command line tools for the analysis of ancient
986 whole-genome duplications. Bioinformatics **35**: 2153-2155
987 **Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, et al.**
988 (2018) Extensive intraspecific gene order and gene structural variations between Mo17
989 and other maize genomes. Nat Genet 50: 1289-1295
990 **Su W, Ou S, Hufford MB, Peterson T** (2021) A Tutorial of EDTA: Extensive De Novo TE
991 Annotator. Methods Mol Biol **2250**: 55-67
992 **Chen S, Zhou Y, Chen Y, Gu J** (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor.
993 Bioinformatics **34**: i884-i890
994 **Kim D, Paggi JM, Park C, Bennett C, Salzberg SL** (2019) Graph-based genome alignment
995 and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol **37**: 907-915

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol **32**: 268-74

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant **13**: 1194-1202.

Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model **61**: 3891-3898.

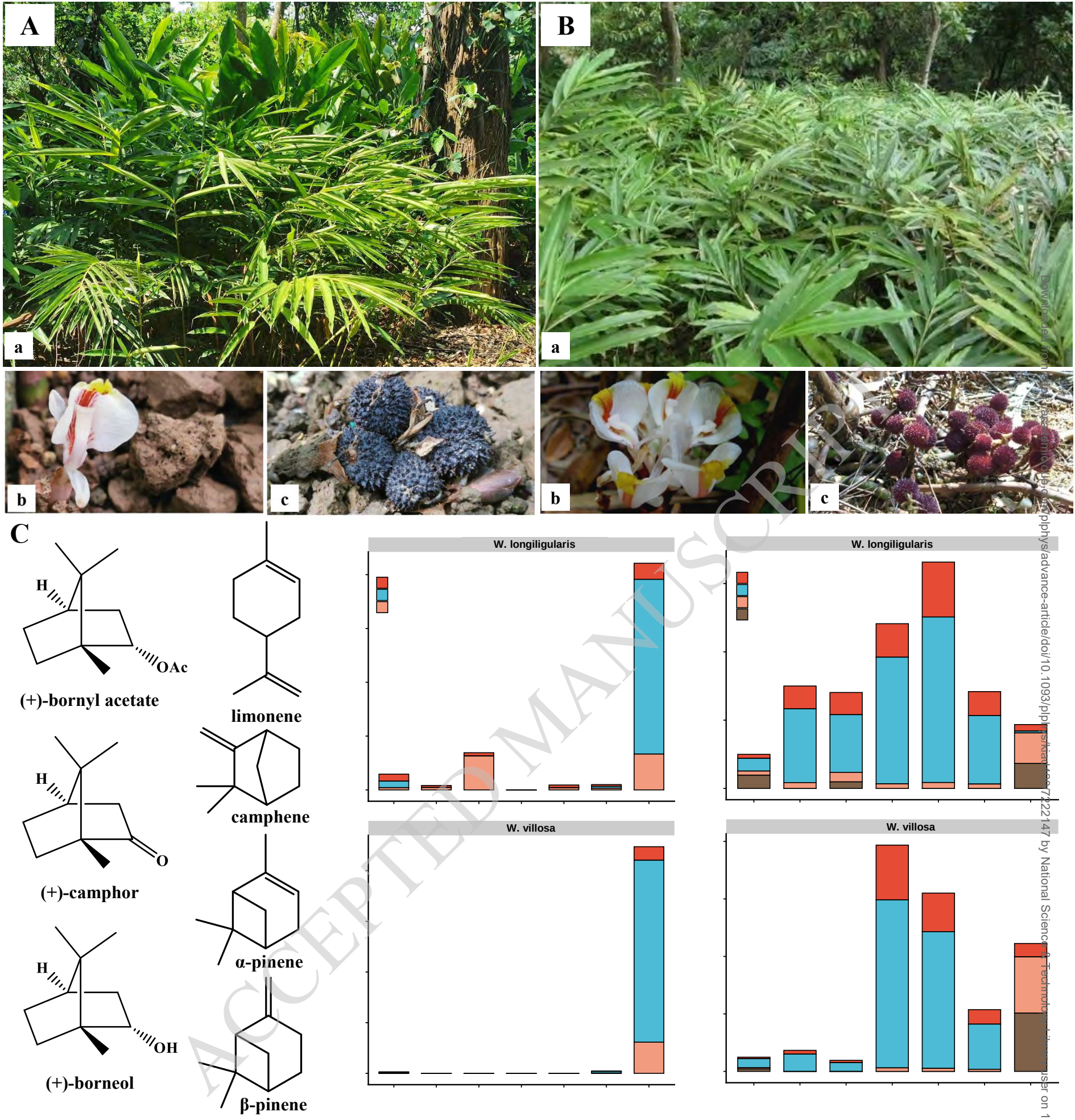

Figure 1 Distribution of volatile terpenoids in *W. longiligularis* and *W. villosa*. (A) Morphological characteristics of *W. longiligularis*. (B) Morphological characteristics of *W. villosa*. a: Plant; b: Flower; c: Fruit. (C) Content distribution of (+)-bornyl acetate, (+)-camphor, (+)-borneol, limonene, camphene, α-pinene, and β-pinene in different tissues of *W. longiligularis* and *W. villosa*.

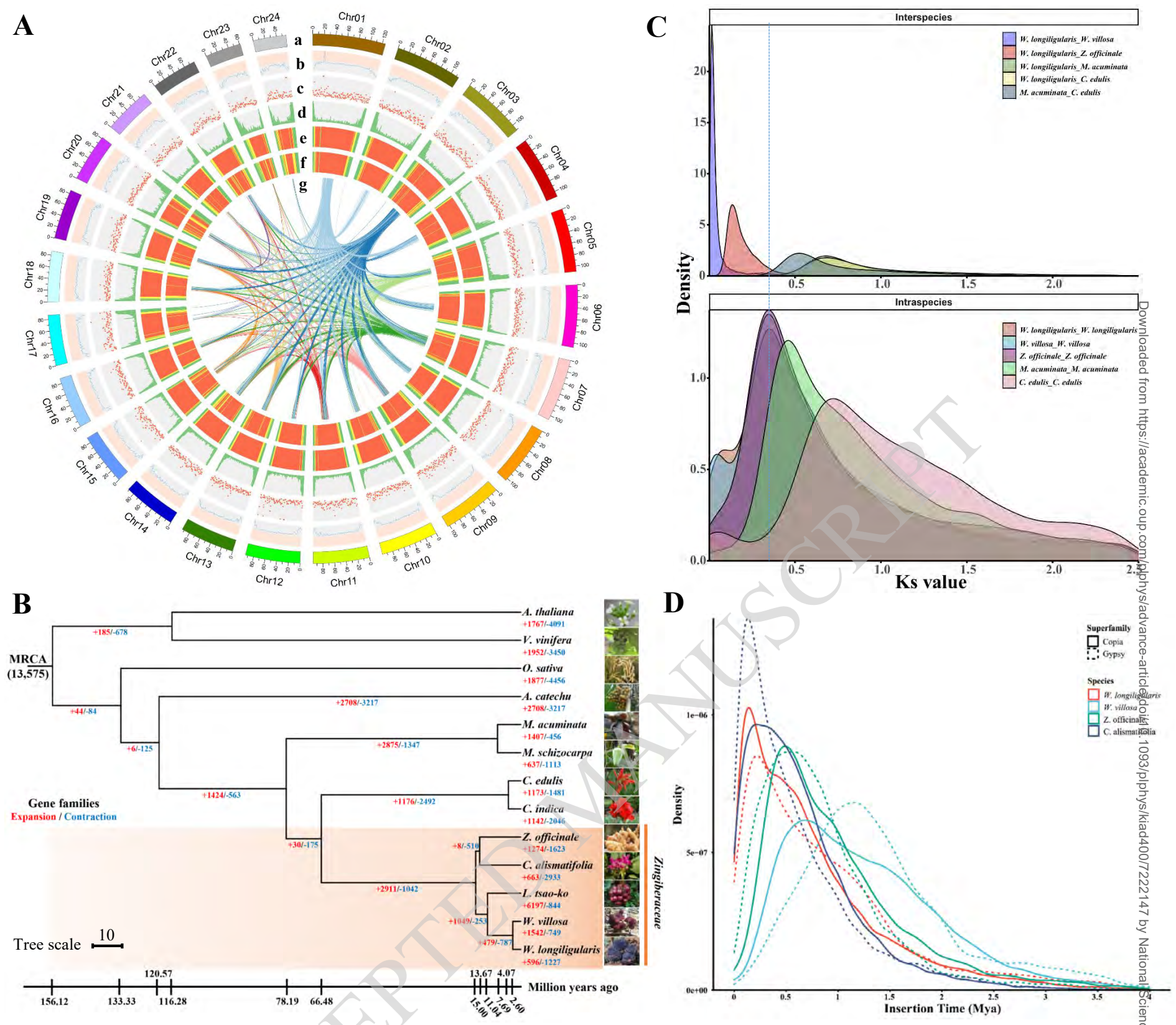

Figure 2 Genomic features of *W. longilingularis*. (A) Circos plot of *W. longilingularis* genome assembly. The window is 1 Mb in size. a, Chromosome karyotypes; b, GC content (min-max, 0-1.0); c, Non-coding RNA (ncRNA) density (min-max, 0-0.002), d, Gene density (min-max, 0-80); e, Transposable elements density (<=0.6, green; 0.6~0.8, yellow; >=0.8, red); f, Long terminal repeat (LTR) density (<=0.6, green; 0.6~0.8, yellow; >=0.8, red); g, Syntenic blocks. (B) Phylogenetic tree and gene family expansions or contractions in 13 species. (C) Distribution of Ks values between *W. longilingularis*, *W. villosa*, *Z. officinale*, *M. acuminata*, and *C. edulis*. (D) Distribution of LTR (*Gypsy* and *Copia*) insertion time of *W. longilingularis*, *W. villosa*, *Z. officinale*, and *C. alismatifolia*.

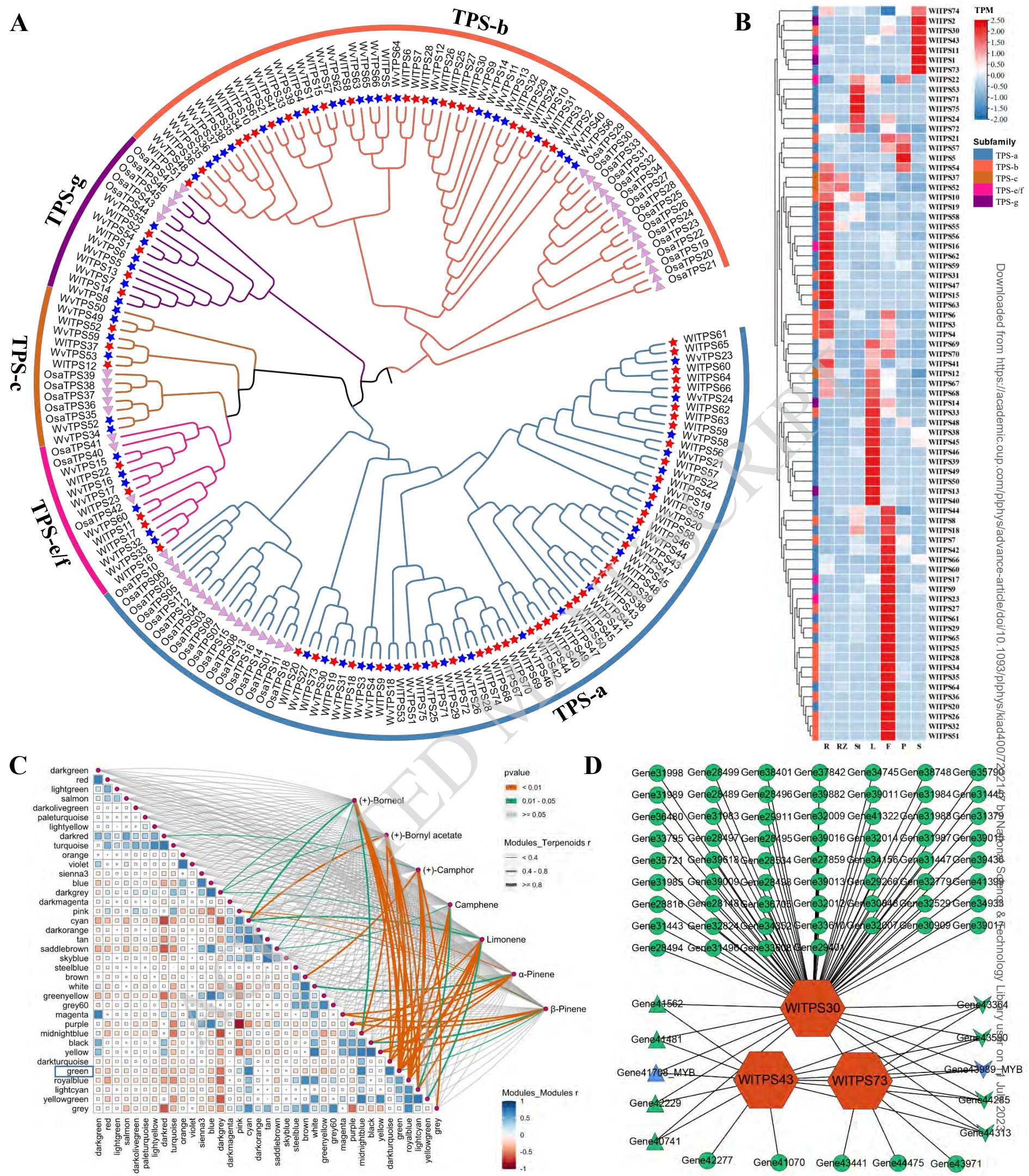

Figure 3 Analysis of TPS gene family in *W. longiligularis*. (A) Phylogenetic tree of TPS genes from *W. longiligularis* (75 genes), *W. villosa* (66 genes), and *O. sativa* (32 genes). Phylogenetic tree ignores branch length information. Red star means *W. longiligularis*, blue star means *W. villosa*, and pink triangle means *O. sativa*. (B) Heatmap (row scale) showing the differential expression of *WlTPSs* according to the transcriptome data from various tissues (R, root; RZ, rhizome; St, stem; L, leaf; F, flower; P, pericarp; S, seeds). (C) Correlation analysis of WGCNA co-expression module with seven terpenoid contents in *W. longiligularis*. (D) Network analysis of TPS-related genes in the green module. The orange hexagon means the TPS genes in the green module.

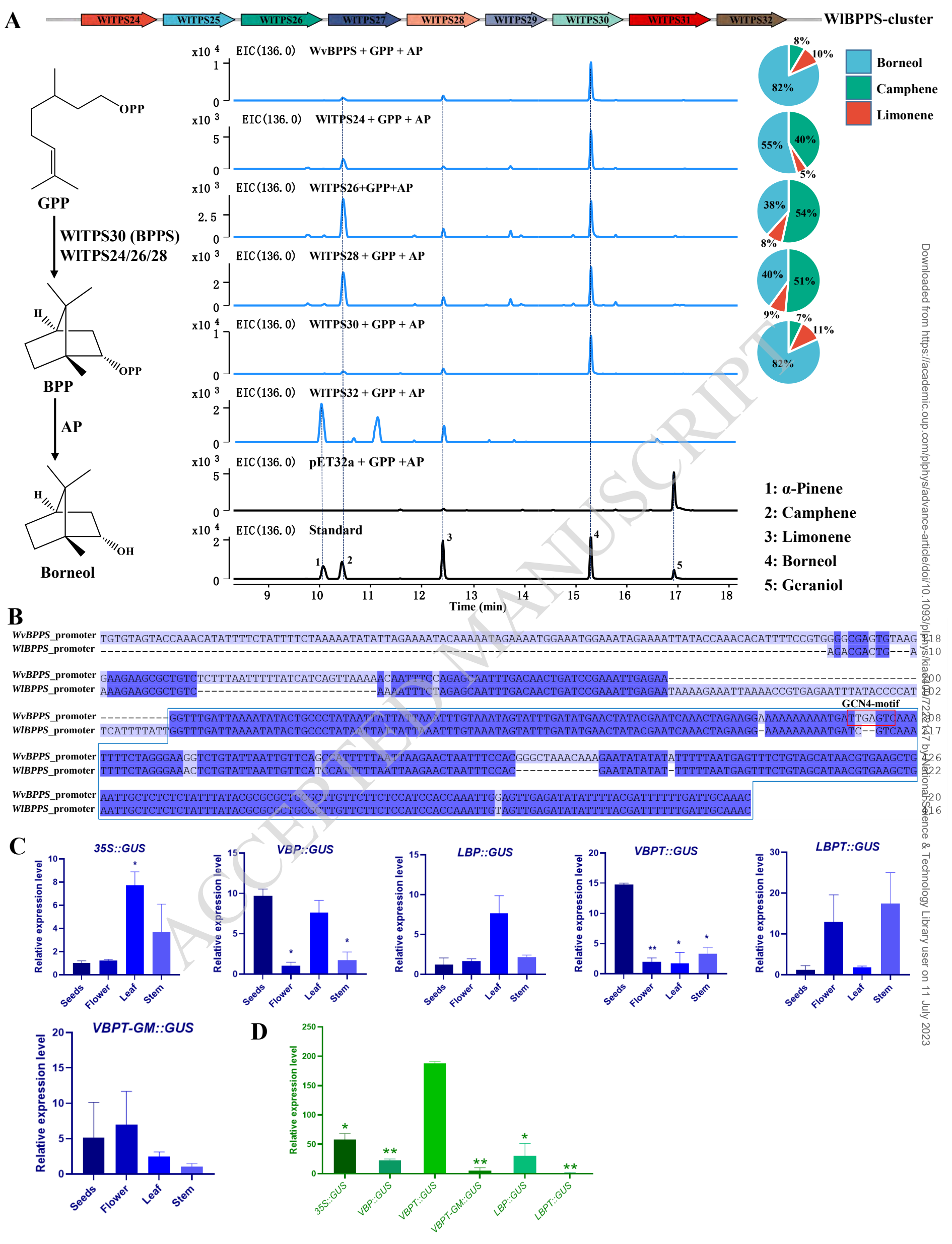

Figure 4. Genetic basis of volatile terpenoids divergence in *W. longiligularis* and *W. villosa*. (A) Functional characterization and comparison of WlBPPS cluster genes in *W. longiligularis*. (B) Multiple sequence alignment of the *WvBPPS* and *WlBPPS* promoters. The blue boxes are the conserved regions of the promoter sequences of the two genes. (C) Analysis of *GUS* gene expression in different transgenic *N. tabacum*. The transgenic *N. tabacum* seeds were statistically analyzed with other tissues (unpaired t-test, *, $P < 0.05$; **, $P < 0.01$). Error bars represent mean ± SD (n = 3 biologically independent samples). (D) Comparison of *GUS* expression in seeds of different transgenic *N. tabacum*. VBPT::GUS compared to other transgenic types (unpaired t-test, *, $P < 0.05$; **, $P < 0.01$). Error bars represent mean ± SD (n = 3 biologically independent samples). VBP::GUS indicates the full-length promoter of *WvBPPS*, VBPT::GUS indicates the truncated to conserved region promoter of *WvBPPS*, VBPT-GM::GUS indicates the truncated to conserved region and GCN4-motif mutant promoter of *WvBPPS*, LBP::GUS indicates the full-length promoter of *WlBPPS* , and LBPT::GUS indicates the truncated to conserved region promoter of *WlBPPS*.

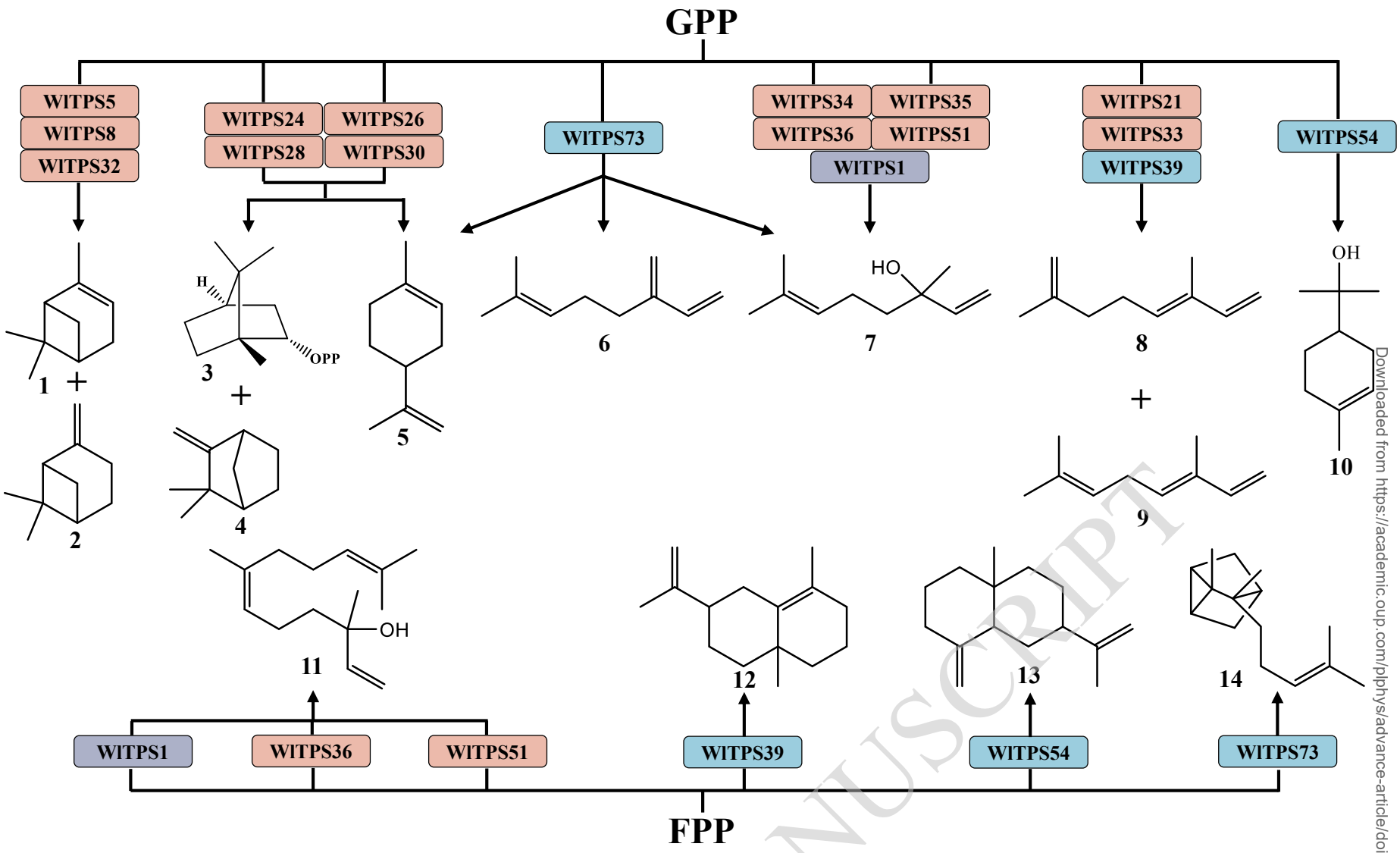

Figure 5 Major products of functional characterization *in vitro* of 17 WlTPSs in *W. longiligularis*. Rectangular background colors depict different subfamilies of WlTPSs: TPS-a (orange), TPS-b (blue), and TPS-g (gray). GPP, geranyl diphosphate; FPP, farnesyl diphosphate; 1, α-pinene; 2, β-pinene; 3, (+)-bornyl diphosphate; 4, camphene; 5, limonene; 6, myrcene; 7, linalool; 8, α-ocimene; 9, β-ocimene; 10, α-terpineol; 11, nerolidol; 12, 4a,8-dimethyl-2-(prop-1-en-2-yl)-1,2,3,4,4a,5,6,7-octahydronaphthalene; 13, β-selinene; 14, α-santalene.

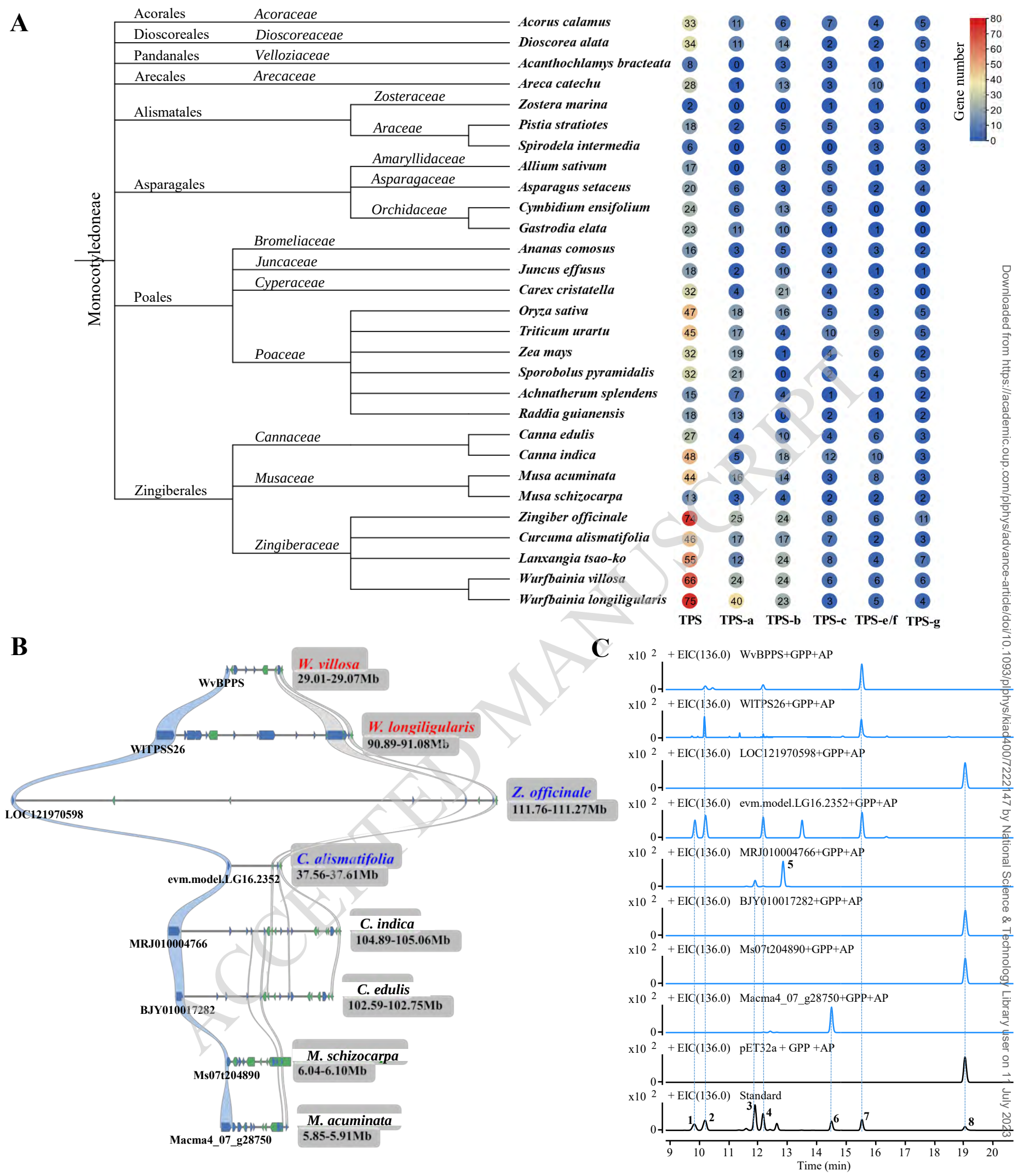

Figure 6 Evolutionary dynamics of BPPS genes and site-directed mutagenesis of WvBPPS. (A) Expansion of TPS genes in *Zingiberaceae*. A total of 916 TPS genes were indentified in 29 monocot plants. Phylogenetic tree ignores branch length information. (B) Synteny analysis showing the conservation of BPPS in Zingiberales species. The blue line shows the syntenic relationship of BPPS among different Zingiberales species. (C) Functional characterization of syntenic genes in Zingiberales species. 1: α-pinene, 2: camphene, 3: terpinolene, 4: limonene, 5: terpinene, 6: α-ocimene, 7: (+)-borneol, 8: geraniol. (D) Relative activity of borneol produced by different plants BPPS and WlBPPS-like. Error bars represent mean ± SD (n = 3 independent experiments). ***P < 0.001; ns, no significant difference; unpaired t-test (n = 3). (E) Relative activity of borneol produced by various mutants. Error bars represent mean ± SD (n = 3 independent experiments). ***P < 0.001; ns, no significant difference; nd, not detected; unpaired t-test (n = 3).

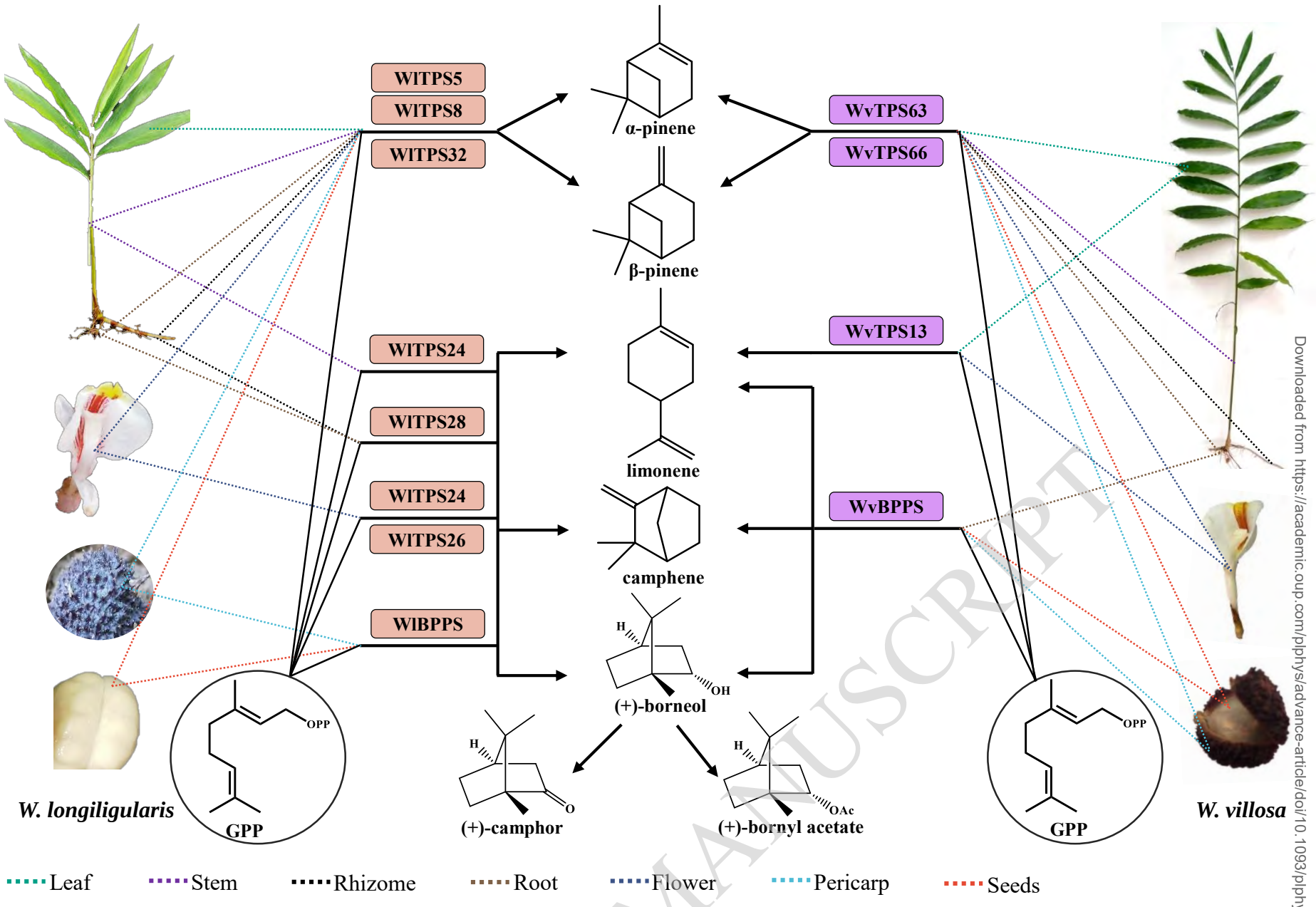

Figure 7 Association analysis between different tissues, main TPSs, and seven main monoterpenoids of the *W. longiligularis* and *W. villosa*. Different color junctions represent TPS genes specific expression tissues.

Parsed Citations

Mei Y, Li L, Fan L, Fan W, Liu L, Zhang F, Hu Z, Wang K, Yang L, Wang Z (2023) The history, stereochemistry, ethnopharmacology and quality assessment of borneol. J Ethnopharmacol 300: 115697
Google Scholar: Author Only Title Only Author and Title

Zhao ZJ, Sun YL, Ruan XF (2023) Bornyl acetate: A promising agent in phytomedicine for inflammation and immune modulation. Phytomedicine 114: 154781
Google Scholar: Author Only Title Only Author and Title

Bahramikia S, Drikvandi P (2022) The abundance of camphor, as a plant essential oil, in various plants and their different parts. Nat Prod Res 3: 1-2
Google Scholar: Author Only Title Only Author and Title

Whittington DA, Wise ML, Urbansky M, Coates RM, Croteau RB, Christianson DW (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. Proc Natl Acad Sci USA 99: 15375-80
Google Scholar: Author Only Title Only Author and Title

Adal AM, Najafianashrafi E, Sarker LS, Mahmoud SS (2023) Cloning, functional characterization and evaluating potential in metabolic engineering for lavender (+)-bornyl diphosphate synthase. Plant Mol Biol 111: 117-130
Google Scholar: Author Only Title Only Author and Title

Hurd MC, Kwon M, Ro DK (2017) Functional identification of a Lippia dulcis bornyl diphosphate synthase that contains a duplicated, inhibitory arginine-rich motif. Biochem Biophys Res Commun 490: 963-968
Google Scholar: Author Only Title Only Author and Title

Wang H, Ma D, Yang J, Deng K, Li M, Ji X, Zhong L, Zhao H (2018) An integrative volatile terpenoid profiling and transcriptomics analysis for gene mining and functional characterization of AvBPPS and AvPS involved in the monoterpenoid biosynthesis in Amomum villosum. Front Plant Sci 9: 846
Google Scholar: Author Only Title Only Author and Title

Yang P, Zhao HY, Wei JS, Zhao YY, Lin XJ, Su J, Li FP, Li M, Ma DM, Tan XK, et al. (2022) Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of Wurfbainia villosa. Plant J 112: 630-645
Google Scholar: Author Only Title Only Author and Title

Doh EJ, Lee G, Jung HJ, Kwon KB, Kim JH (2020) Chemotaxonomic monitoring of genetically authenticated Amomi Fructus using high-performance liquid chromatography-diode array detector with chemometric analysis. Molecules 25: 4581
Google Scholar: Author Only Title Only Author and Title

Ma Y, Cui G, Chen T, Ma X, Wang R, Jin B, Yang J, Kang L, Tang J, Lai C, et al. (2021) Expansion within the CYP71D subfamily drives the heterocyclization of tanshinones synthesis in Salvia miltiorrhiza. Nat Commun 12: 685
Google Scholar: Author Only Title Only Author and Title

Zhong C, Chen C, Gao X, Tan C, Bai H, Ning K (2022) Multi-omics profiling reveals comprehensive microbe-plant-metabolite regulation patterns for medicinal plant Glycyrrhiza uralensis Fisch. Plant Biotechnol J 20: 1874-1887
Google Scholar: Author Only Title Only Author and Title

Jia Q, Brown R, Köllner TG, Fu J, Chen X, Wong GK, Gershenzon J, Peters RJ, Chen F (2022) Origin and early evolution of the plant terpene synthase family. Proc Natl Acad Sci USA 119: e2100361119

Google Scholar: Author Only Title Only Author and Title

Bao T, Kimani S, Li Y, Li H, Yang S, Zhang J, Wang Q, Wang Z, Ning G, Wang L, et al. (2023) Allelic Variation of Terpene Synthases Drives Terpene Diversity in the Wild Species of the Freesia Genus. Plant Physiol 18: kiad172
Google Scholar: Author Only Title Only Author and Title

Rosenkranz M, Chen Y, Zhu P, Vlot AC (2021) Volatile terpenes-mediators of plant-to-plant communication. Plant J 108: 617-631
Google Scholar: Author Only Title Only Author and Title

Nagegowda DA, Gupta P (2020) Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. Plant Sci 294: 110457
Google Scholar: Author Only Title Only Author and Title

Guo X, Fang D, Sahu SK, Yang S, Guang X, Folk R, Smith SA, Chanderbali AS, Chen S, Liu M, et al. (2021) Chloranthus genome provides insights into the early diversification of angiosperms. Nat Commun 12: 6930
Google Scholar: Author Only Title Only Author and Title

Barbosa GB, Jayasinghe NS, Natera SHA, Inutan ED, Peteros NP, Roessner U (2017) From common to rare Zingiberaceae plants-A metabolomics study using GC-MS. Phytochemistry 140: 141-150
Google Scholar: Author Only Title Only Author and Title

Tunnisa F, Nur Faridah D, Afriyanti A, Rosalina D, Ana Syabana M, Darmawan N, Dewi Yuliana N (2022) Antioxidant and antidiabetic compounds identification in several indonesian underutilized Zingiberaceae spices using SPME-GC/MS-based volatilomics and in silico methods. Food Chem X 14: 100285
Google Scholar: Author Only Title Only Author and Title

Kulyal P, Acharya S, Ankari AB, Kokkiripati PK, Tetali SD, Raghavendra AS (2021) Variable secondary metabolite profiles across cultivars of Curcuma longa L. and C. aromatica Salisb. Front Pharmacol 12: 659546
Google Scholar: Author Only Title Only Author and Title

Peng W, Li P, Ling R, Wang Z, Feng X, Liu J, Yang Q, Yan J (2022) Diversity of volatile compounds in ten varieties of Zingiberaceae. Molecules 27: 565
Google Scholar: Author Only Title Only Author and Title

Ivanović M, Makoter K, Razboršek M (2021) Comparative study of chemical composition and antioxidant activity of essential oils and crude extracts of four characteristic Zingiberaceae herbs. Plants (Basel) 10: 501
Google Scholar: Author Only Title Only Author and Title

Li HL, Wu L, Dong Z, Jiang Y, Jiang S, Xing H, Li Q, Liu G, Tian S, Wu Z, et al. (2021) Haplotype-resolved genome of diploid ginger (Zingiber officinale) and its unique gingerol biosynthetic pathway. Hortic Res 8: 189
Google Scholar: Author Only Title Only Author and Title

Liao X, Ye Y, Zhang X, Peng D, Hou M, Fu G, Tan J, Zhao J, Jiang R, Xu Y, et al. (2022) The genomic and bulked segregant analysis of Curcuma alismatifolia revealed its diverse bract pigmentation. Abiotech 3: 178-196
Google Scholar: Author Only Title Only Author and Title

Yin Y, Xie X, Zhou L, Yin X, Guo S, Zhou X, Li Q, Shi X, Peng C, Gao J (2022) A chromosome-scale genome assembly of turmeric provides insights into curcumin biosynthesis and tuber formation mechanism. Front Plant Sci 13: 1003835
Google Scholar: Author Only Title Only Author and Title

Li P, Bai G, He J, Liu B, Long J, Morcol T, Peng W, Quan F, Luan X, Wang Z, et al. (2022) Chromosome-level genome assembly of Amomum tsao-ko provides insights into the biosynthesis of flavor compounds. Hortic Res 9: uhac211
Google Scholar: Author Only Title Only Author and Title

Lin XJ, Liang HL, Zhao HY, Wu QW, Huang LX, Wu SR, Yang JF (2022) Comparison of BPPS promoters between Amomum villosum and Amomum longiligulare and identification of GCN4 motif positive regulation. Chinese Traditional and Herbal Drugs 53: 6159-6166
Google Scholar: Author Only Title Only Author and Title

Ao H, Wang J, Chen L, Li S, Dai C (2019) Comparison of Volatile Oil between the Fruits of Amomum villosum Lour. and Amomum villosum Lour. var. xanthioides T. L. Wu et Senjen Based on GC-MS and Chemometric Techniques. Molecules 24: 1663
Google Scholar: Author Only Title Only Author and Title

Zhao H, Li M, Zhao Y, Lin X, Liang H, Wei J, Wei W, Ma D, Zhou Z, Yang J (2021) A Comparison of Two Monoterpenoid Synthases Reveals Molecular Mechanisms Associated with the Difference of Bioactive Monoterpenoids Between Amomum villosum and Amomum longiligulare. Front Plant Sci 12: 695551
Google Scholar: Author Only Title Only Author and Title

He J, Verstappen F, Jiao A, Dicke M, Bouwmeester HJ, Kappers IF (2022) Terpene synthases in cucumber (Cucumis sativus) and their contribution to herbivore-induced volatile terpenoid emission. New Phytol 233: 862-877
Google Scholar: Author Only Title Only Author and Title

Li J, Wang Y, Dong Y, Zhang W, Wang D, Bai H, Li K, Li H, Shi L (2021) The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. Hortic Res 8: 53
Google Scholar: Author Only Title Only Author and Title

Wang X, Gao Y, Wu X, Wen X, Li D, Zhou H, Li Z, Liu B, Wei J, Chen F, et al. (2021) High-quality evergreen azalea genome reveals tandem duplication-facilitated low-altitude adaptability and floral scent evolution. Plant Biotechnol J 19: 2544-2560
Google Scholar: Author Only Title Only Author and Title

Wu C, Washida H, Onodera Y, Harada K, Takaiwa F (2000) Quantitative nature of the Prolamin-box, ACGT and AACA motifs in a rice glutelin gene promoter: minimal cis-element requirements for endosperm-specific gene expression. Plant J 23: 415-21
Google Scholar: Author Only Title Only Author and Title

Kim JS, Chae S, Jun KM, Pahk YM, Lee TH, Chung PJ, Kim YK, Nahm BH (2017) Genome-wide identification of grain filling genes regulated by the OsSMF1 transcription factor in rice. Rice 10: 16
Google Scholar: Author Only Title Only Author and Title

Onodera Y, Suzuki A, Wu CY, Washida H, Takaiwa F (2001) A rice functional transcriptional activator, RISBZ1, responsible for endosperm-specific expression of storage protein genes through GCN4 motif. J Biol Chem 276: 14139-52
Google Scholar: Author Only Title Only Author and Title

Timalsina D, Devkota HP (2021) Eclipta prostrata (L.) L. (Asteraceae): ethnomedicinal uses, chemical constituents, and biological activities. Biomolecules 11: 1738
Google Scholar: Author Only Title Only Author and Title

Mohammadhosseini M, Frezza C, Venditti A, Mahdavi B (2022) An overview of the genus Aloysia Paláu (Verbenaceae): Essential oil composition, ethnobotany and biological activities. Nat Prod Res 36: 5091-5107
Google Scholar: Author Only Title Only Author and Title

Cheng KK, Nadri MH, Othman NZ, Rashid SNAA, Lim YC, Leong HY (2022) Phytochemistry, Bioactivities and Traditional Uses of Michelia × alba. Molecules 27: 3450
Google Scholar: Author Only Title Only Author and Title

Karpiński TM (2020) Essential Oils of Lamiaceae Family Plants as Antifungals. Biomolecules 10: 103
Google Scholar: Author Only Title Only Author and Title

Wang J, Su B, Jiang H, Cui N, Yu Z, Yang Y, Sun Y (2020) Traditional uses, phytochemistry and pharmacological activities of the genus Cinnamomum (Lauraceae): A review. Fitoterapia 146: 104675
Google Scholar: Author Only Title Only Author and Title

Tian Z, Zeng P, Lu X, Zhou T, Han Y, Peng Y, Xiao Y, Zhou B, Liu X, Zhang Y, et al. (2022) Thirteen Dipterocarpoideae genomes provide insights into their evolution and borneol biosynthesis. Plant Commun 3: 100464
Google Scholar: Author Only Title Only Author and Title

Czechowski T, Branigan C, Rae A, Rathbone D, Larson TR, Harvey D, Catania TM, Zhang D, Li Y, Salmon M, et al. (2022) Artemisia annua L. plants lacking Bornyl diPhosphate Synthase reallocate carbon from monoterpenes to sesquiterpenes except artemisinin. Front Plant Sci 13: 1000819
Google Scholar: Author Only Title Only Author and Title

Ma R, Su P, Guo J, Jin B, Ma Q, Zhang H, Chen L, Mao L, Tian M, Lai C, et al. (2021) Bornyl Diphosphate Synthase from Cinnamomum burmanni and its application for (+)-borneol biosynthesis in yeast. Front Bioeng Biotechnol 9: 631863
Google Scholar: Author Only Title Only Author and Title

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun 11: 1432
Google Scholar: Author Only Title Only Author and Title

Cheng H, Concepcion GT, Feng X, Zhang H, Li H (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods 18: 170-175
Google Scholar: Author Only Title Only Author and Title

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL (2016) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst 3: 95-8
Google Scholar: Author Only Title Only Author and Title

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP et al. (2017) De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356: 92-95
Google Scholar: Author Only Title Only Author and Title

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst 3: 99-101
Google Scholar: Author Only Title Only Author and Title

Seppey M, Manni M, Zdobnov EM (2019) BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol 1962: 227-245
Google Scholar: Author Only Title Only Author and Title

Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33: W465-7.
Google Scholar: Author Only Title Only Author and Title

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29: 2933-

5

Google Scholar: Author Only Title Only Author and Title

Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, et al. (2021) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res 49: D192-D200

Google Scholar: Author Only Title Only Author and Title

Chan PP, Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962: 1-14

Google Scholar: Author Only Title Only Author and Title

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20: 238

Google Scholar: Author Only Title Only Author and Title

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-3

Google Scholar: Author Only Title Only Author and Title

Xu B, Yang Z (2013) PAMLX: a graphical user interface for PAML. Mol Biol Evol 30: 2723-4

Google Scholar: Author Only Title Only Author and Title

Mendes FK, Vanderpool D, Fulton B, Hahn MW (2020) CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics btaa1022

Google Scholar: Author Only Title Only Author and Title

Zwaenepoel A, Van Y (2019) wgd-simple command line tools for the analysis of ancient whole-genome duplications. Bioinformatics 35: 2153-2155

Google Scholar: Author Only Title Only Author and Title

Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, et al. (2018) Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nat Genet 50: 1289-1295

Google Scholar: Author Only Title Only Author and Title

Su W, Ou S, Hufford MB, Peterson T (2021) A Tutorial of EDTA: Extensive De Novo TE Annotator. Methods Mol Biol 2250: 55-67

Google Scholar: Author Only Title Only Author and Title

Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884-i890

Google Scholar: Author Only Title Only Author and Title

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37: 907-915

Google Scholar: Author Only Title Only Author and Title

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32: 268-74

Google Scholar: Author Only Title Only Author and Title

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant 13: 1194-1202.

Google Scholar: Author Only Title Only Author and Title

Eberhardt J, Santos-Martins D, Tillack AF, Forli S (2021) AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. J Chem Inf Model 61: 3891-3898.

Google Scholar: Author Only Title Only Author and Title