Food Chemistry: Molecular Sciences 5 (2022) 100129

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

journal homepage: www.sciencedirect.com/journal/food-chemistry-molecular-sciences

Transcriptome sequencing and gas chromatography–mass spectrometry analyses provide insights into β-caryophyllene biosynthesis in *Brassica campestris*

Haibin Wang [a,b,c], Chen Zong [a,b,c], Aimei Bai [a,b,c], Shuilin Yuan [a,b,c], Yan Li [a,b,c], Zhanghong Yu [a,b,c], Ruiping Tian [a], Tongkun Liu [a,b,c], Xilin Hou [a,b,c], Ying Li [a,b,c,*]

[a] National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, Jiangsu Province, China
[b] Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (East China), Ministry of Agriculture and Rural Affairs of China, China
[c] College of Horticulture, Nanjing Agricultural University, Nanjing 210095, Jiangsu Province, China

ARTICLE INFO

Keywords:
Non-heading Chinese cabbage
β-Caryophyllene
HS-SPME-GC/MS
Transcriptome analysis

ABSTRACT

Sesquiterpenes are important defensive secondary metabolites and aroma components. However, limited information is available on the mechanism of sesquiterpene formation and composition in the non-heading Chinese cabbage (NHCC) leaf. Therefore, headspace solid-phase microextraction/gas chromatography–mass spectrometry (HS-SPME/GC–MS) combined with transcriptome analysis was used to study the mechanism of volatile organic compound formation. A total of 26 volatile organic compounds were identified in two NHCC cultivars 'SZQ' and 'XQC' and their F1 hybrids. Among these, sesquiterpene β-caryophyllene was identified only in 'XQC' and F1. Five genes encoding caryophyllene synthase were identified. The candidate β-caryophyllene synthase genes *BcTPSa11* and *BcTPSa21* had high expression levels only in 'XQC' and F1. In addition, several transcription factors of MYB-related, MYB, bHLH, and AP2/ERF families were identified by co-expression, suggesting that they regulate β-caryophyllene biosynthesis. Our results provide a molecular basis for sesquiterpene biosynthesis as well as insights into the regulatory network of β-caryophyllene in NHCC.

1. Introduction

Non-heading Chinese cabbage (NHCC) is an economically important leafy vegetable with abundant germplasm resources in China, Japan, and Korea (Chen et al., 2016). It is consumed by people in large amounts because of its attractive unique flavor. In addition, the aroma of NHCC is a key feature to evaluate its quality and is an important trait for breeding. Sesquiterpene derivatives such as artemisinin, β-caryophyllene (BCP), cedrol, and nerolidol are the main volatile terpenes in plants, and each contributes a distinct aroma. They are found in scented fruits, vegetables, and flowers, where they perform numerous key functions, particularly in protection against biotic and abiotic stresses. Thus, exploring the sesquiterpene components and regulatory mechanisms in NHCC is meaningful.

Terpenes—monoterpenes, sesquiterpenes, and diterpenes—constitute an important class of secondary metabolites in plants. A series of genes involved in the terpenoid synthesis pathway have been identified in plants. The terpenoid synthase (TPS) gene, a key gene in terpene synthesis, is responsible for the characteristic aroma in plants. The TPS gene family has been explored in *Arabidopsis thaliana*, *Vitis vinifera*, *Populus trichocarpa*, *Oryza sativa*, *Solanum lycopersicum*, and *Glycine* max (Aubourg, Lecharny, & Bohlmann, 2002; Chen et al., 2014; Falara et al., 2011; Irmisch, Jiang, Chen, Gershenzon, & Köllner, 2014; Liu et al., 2014; Martin et al., 2010). The most prominent feature of TPS proteins is the aspartic and arginine acid conserved domain, which comprises the highly conserved signature domains DDxxD and RRX8w. According to the similarity between the known genome sequence and the amino acid sequence of terpene synthase, the plant TPS gene family can be divided into TPSa–g (seven subfamilies), and sesquiterpene synthase is mainly concentrated in the TPSa subfamily (Bohlmann, Meyer-Gauen, & Croteau, 1998). However, not much information is available on the expression of TPSa subfamily genes or sesquiterpene synthesis and regulation in NHCC.

The functional diversity and molecular evolution of the TPSa subfamily underlie the chemical diversity of bioactive sesquiterpenes in plants. *CsTPS1*, which is expressed specifically during fruit maturation

* Corresponding author.
E-mail address: yingli@njau.edu.cn (Y. Li).

https://doi.org/10.1016/j.fochms.2022.100129
Received 1 July 2022; Received in revised form 28 July 2022; Accepted 6 August 2022
Available online 13 August 2022

in citrus, is a key gene involved in the production of the sesquiterpene aroma compound valencene (Sharon-Asa et al., 2003). Two sesquiterpene synthase genes, *SgSTPS1* and *SgSTPS2* expressing BCP and β-copaene, respectively, were identified in the *Sindora glabra* transcriptome (Yu, Yang, Yin, Li, & Zou, 2018). However, the genes responsible for producing the sesquiterpene BCP in NHCC remain unknown. Thus, the identification and characterization of TPSa genes were considered in this study.

Jasmonate and its derivative methyl jasmonate (MeJA) modulate sesquiterpene accumulation. The α-copaene and BCP levels significantly increased after MeJA treatment and might enhance endogenous gibberellin biosynthesis, thereby regulating sesquiterpene synthesis (Yu, Chen, Yang, Li, & Zou, 2021). Moreover, transcription factors (TFs) play an important role in regulating sesquiterpene biosynthesis. Overexpression of the transcription factor HAC1 improves nerolidol production in engineered yeast (Qu et al., 2020). MYB21 interacts with MYC2 to control terpene synthase gene expression in the flowers of *Freesia hybrida* and *A. thaliana* (Yang et al., 2020). The bHLH transcription factor MYC2 can directly bind to the promoters of *TPS21* and *TPS11* genes to increase sesquiterpene content, especially BCP in *A. thaliana* (Hong, Xue, Mao, Wang, & Chen, 2012). Thus, the potential regulatory networks of sesquiterpene biosynthesis in NHCC were evaluated by co-expression analysis in this study.

Despite these advances, the understanding of the regulatory mechanisms and functioning of sesquiterpenes is still limited in NHCC. Thus, the main objective of this study was to investigate the concentration of volatile constituents and the expression profiles of the related genes in NHCC, with the goal of deciphering the molecular mechanism

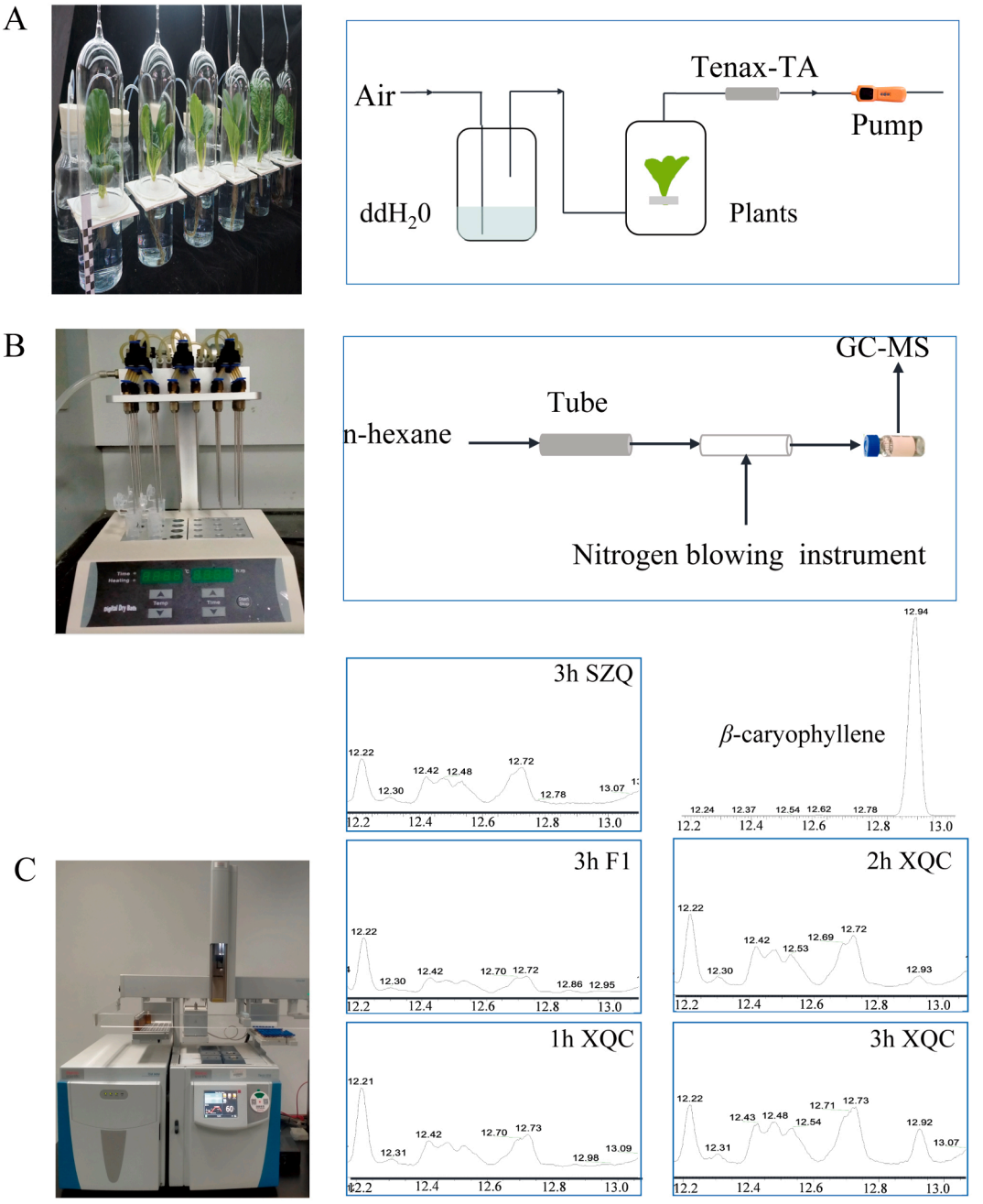


Fig. 1. Collection of released volatiles and GC–MS analysis of *Brassica campestris*. (A) In-kind diagram and principal diagram of a collection of released volatiles. (B) In-kind diagram and principal diagram of desorption and enrichment device with nitrogen blowing instrument. (C) Gas chromatography–mass spectrometry (GC–MS) system and GC–MS analysis of *B. campestris*. The collection tube was loaded with Tenax (200 mg) for real-time collection during GC–MS (n = 1).

underlying sesquiterpene biosynthesis and regulation. Therefore, the roles of *BcTPSa* subfamily members, gene taxonomy, chromosomal location, and phylogenetic relationship of NHCC were investigated using bioinformatics methods. Moreover, headspace solid-phase microextraction/gas chromatography–mass spectrometry (HS-SPME/GC–MS) was used to investigate the changes in the volatile constituents, and RNA-seq was used to probe the expression of genes involved in terpenoid metabolism during NHCC development. Comparisons of volatile constituents and gene expression were performed at different developmental stages from three materials, including 'XQC', 'SZQ', and F1.

2. Materials and methods

2.1. Plant materials

Brassica campestris F1 hybrids and parent lines ('XQC' and 'SZQ') were obtained from the greenhouse at Nanjing Agricultural University, China. Leaf samples of different materials were collected after 100-μM MeJA treatment. In addition, 'XQC' and 'SZQ' samples were added at 20 and 40 days to observe differences in volatile organic compounds (VOCs) at different stages of development **(Fig. S1)**. After surface cleaning with ddH_2O, 1.5 g of leaves was immediately frozen in liquid nitrogen and stored at −80 °C until use. Three biological replicates were used for GC–MS analysis and RNA-seq analysis at each time point. Each replicate comprised at least six leaves collected from six plants mixed in equal amounts.

2.2. Collection and analysis of dynamic headspace VOCs

We collected and analyzed volatiles from *B. campestris* using HS-SPME/GC–MS **(**Fig. 1**)**. The released volatiles were collected **(**Fig. 1**A)** and passed through a desorption and enrichment device using a nitrogen blowing instrument **(**Fig. 1**B)**. The end of the Tenax tube was connected to an air pump set at a flow rate of 500 mL/min for 1, 2, and 3 h for 'XQC', and for 3 h for 'SZQ' and 'F1′. The sample was eluted with 3 mL of *n*-hexane and concentrated to 500 μl with a nitrogen generator (QPN-15LP) after elution and stored at − 20 °C until use. Helium was used as the carrier gas at a linear velocity of 1.5 mL/min. The initial oven temperature was held at 60 °C for 2 min, increased to 140 °C at a rate of 10 °C/min and held for 6 min, and the temperature was then increased to 250 °C at a rate of 20 °C/min and held for 1 min.

2.3. Sample preparation and GC–MS analysis

To further explore the types of VOCs in *B. campestris*, 1.5 g of fresh leaves was ground to a powder in liquid nitrogen. VOCs of leaves were detected by a gas chromatograph (TRACE 1310, Thermo Scientific), coupled to a triple quadrupole mass spectrometer based on the TSQ 9000 Thermo Scientific platform at 250 °C for 3 min in split mode (flow rate 1 mL/min). Three replicates of each assay were performed. Subsequently, 1.5 g of the powder was transferred to a 20-mL headspace vial (Agilent, CA, USA) containing 2-mL NaCl saturated solution to inhibit any enzyme reaction. An 85-μm CAR/PDMS fiber was exposed to the headspace of the sample for 40 min at 35 °C for solid-phase microextraction (SPME) analysis.

Helium was used as the carrier gas at a linear velocity of 1.0 mL/min. The initial oven temperature was held at 35 °C for 3 min, and the temperature was gradually increased to 45 °C at a rate of 8 °C/min, 140 °C at a rate of 6 °C/min, and 230 °C at a rate of 10 °C/min and held for 6 mins. The injector temperature was kept at 200 °C, and the detector temperature was 250 °C. MS was operated under an electron impact ionization mode of 70 eV with scanning in the range *m/z* 30–500 amu at 1 s intervals. The quadrupole mass detector, ion source, and transfer line temperatures were set at 200 °C, 250 °C, and 270 °C, respectively. Volatile compounds were identified by comparing the mass spectra with the National Institute of Standards and Technology (NIST) and linear retention index.

2.4. RNA extraction and RNA-seq analysis

Total RNA was extracted from *B. campestris* using a plant RNA isolation kit (Mage, Beijing, China), according to the manufacturer's instructions, for RNA library preparation and sequencing.

RNA integrity and concentration were assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). After the total RNA samples were verified, the MGIEasy RNA library preparation kit was used for library construction. Oligo-dT magnetic beads were used for enrichment and purification. The fragment buffer was added to purified mRNA to shorten its fragments, and fragmented mRNA was used as a template for reverse transcription into double-stranded cDNA with six base random primers, and the product was purified. The purified double-stranded cDNA was subjected to end-repair, tailing, adaptor ligation, and PCR amplification. After verification, the amplified product was denatured, single-stranded, and cyclized, and the cyclized product was inspected. The DNB chip was inserted into the DNBSEQ-T7 system (MGI Tech Co., ltd., Shenzhen, China) for sequencing, and 150-bp double-ended sequencing reads were obtained. Three biological replicate samples from each time point were used for library construction. Raw reads were obtained from each sample after sequencing. After removing primer sequences, adaptor sequences, and low-quality reads, clean reads consisted of more than 80 % base pairs with Q-value $\geq$ 30. High-quality reads were compared with the spliced transcriptome. The transcripts per million (TPM) value was calculated for each transcription region to quantify its expression abundance and variations, and differentially expressed genes (DEGs) were identified using DESeq2 (1.20.0). The false discovery rate (FDR) was controlled by the Benjamini–Hochberg method to correct the hypothesis test probability (p-value). The screening criteria for DEGs were $|\log2FC| \geq 1$ and FDR < 0.05.

2.5. Identification of candidate genes and transcription factors

In order to identify the structural genes related to the synthesis of BCP, the differential structural genes of terpene synthesis pathway were identified using DESeq2 (1.20.0). To search for the transcription factors that were co-expressed with candidate genes, Pearson's correlation analysis between the expression level of all the differential transcription factor and candidate genes was performed using OmicShare Tools. The transcription factors with Pearson's correlation coefficients $\geq$ 0.9 were selected for further analysis.

2.6. Genome-wide identification and phylogenetic analysis of BcTPSa gene

Firstly, the Arabidopsis TPS protein and other species protein sequences were retrieved and downloaded from the Ensemble Plants database (https://plants.ensembl.org/index.html), meanwhile the *Raphanus sativus* protein sequences were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). The Arabidopsis TPS proteins were used as the queries to perform a BLASTP search against the local protein database of NHCC001 and other species using bioedit software. All the sequences obtained from the alignment were placed on the CDD search in the NCBI website for conservative domain analysis, and the protein sequences containing both pf03936 and pf01397 domains were selected to obtain the final protein sequence containing TPS domain. The molecular weight (Mw), isoelectric point (pI), instability index (II), aliphatic index (AI), and grand average of hydropathicity (GRAVY) of these identified proteins were investigated using ExPASy (https://web.expasy.org/protparam/) online software. Subcellular localization of them were predicted based on the CELLO v2.5 software (http://cello.life.nctu.edu.tw/).

To understand the evolutionary relationship of NHCC001 TPSa genes, multiple sequence alignments of the identified TPS proteins of NHCC001 and other species, were performed using ClustalW tool and the maximum likelihood (ML) tree was constructed by Tbtools software with bootstrap value of 1000. The remaining 7 species were classified according to the TPS classification of Arabidopsis, and all TPSa were combined with caryophyllene synthase recorded by NCBI to construct the evolutionary tree again. The chromosomal location and exon–intron structures of non-heading Chinese cabbage TPSa genes were obtained from the genome annotation files. The chromosome physical location of the TPSa genes and the exon–intron structure were displayed using the Tbtools v1.09.

In order to further understand the phylogenetic relationship of BCP genes, multiple sequence alignment of 29 BCP members was conducted. Amino acid sequence alignment was conducted with MUSCLE and phylogenetic analysis was done. Maximum likelihood evolutionary tree by Tbtools software with 1000 bootstrap replicates.

2.7. Subcellular localization and luciferase reporter experiment

The cDNA libraries of *Brassica campestris* were prepared using the Reverse Transcription Kit (PrimeScriptTM RT Master Mix, Takara) with random primers. The complete ORF sequence of *BcMYC2*, *BcMYB122* and *BcERF23* without a stop codon were subcloned to PJX003 to generate *35S::BcMYC2*, *35S::BcMYB122* and *35S::BcERF23* effector constructs. The agrobacterium solution containing PJX003-BcMYC2-GFP, PJX003-BcMYB122-GFP, PJX003-BcERF23-GFP was injected into tobacco epidermal cells. At 3 days after injection, green fluorescent protein (GFP) fluorescence was observed via confocal laser scanning microscopy (LSM780, Zeiss). We constructed the putative promoter regions Pro-*BcTPSa21*-300 and Pro-*BcTPSa21*-1050 of *BcTPSa21* promoters to pGreenII 0800-LUC vector through *Xho*I and *Kpn*I to create reporter constructs. Agrobacterium cells containing reporter construct and TF were injected into tobacco leaves at a ratio of 1:2. In addition, Agrobacterium cells containing reporter construct and empty vector were injected at the same ratio as a control.

2.8. Quantitative real-time RT-PCR

Total RNA used for qRT-PCR was extracted using HiPure Plant RNA Mini Kit (R4151-02, Mage). Then the RNA was reverse transcribed to the first-strand cDNA using the Reverse Transcription Kit (PrimeScriptTM RT Master Mix, Takara). The qRT-PCR was performed with a CFX96 (Bio-rad, USA) and the SYBR qPCR Master Mix kit (Vazyme). The relative expression level was calculated according to the $2^{-\triangle\triangle Ct}$ method. The Tubulin gene was used as a reference gene. Ct represents the PCR cycle number at which the amount of target reaches a fixed threshold.

3. Results

3.1. Release of BCP from B. campestris leaves

To analyze the components and release pattern of aroma with specific flavor in *B. campestris*, we used a collection system and a desorption and enrichment device with a nitrogen blowing instrument combined with a GC–MS system (Fig. 1). Only one volatile substance, BCP, was determined. BCP content increased with collection time in the 'XQC' cultivar, whereas it was undetectable in the 'SZQ' cultivar and F1 even after 3 h of continuous collection (Fig. 1**C**).

3.2. VOCs in B. campestris

In this study, F1 hybrids were generated by crossing 'XQC' and 'SZQ'. The flavor-related trait, BCP, could be inherited by F1 hybrids (Fig. 2). A total of 26 VOCs were identified in 'XQC', 'SZQ', and F1 leaves by HS-SPME/GC–MS (**Table. S0**). GC–MS analysis of VOCs in 'XQC' leaves at 20, 40, and 60 days showed that BCP was only detectable at 60 days (Fig. 2**A**). BCP content significantly increased in 'XQC' and F1 after 100-μM MeJA treatment at 0, 6, and 12 h (Fig. 2**B, C, and E**). Similar to our results, jasmonate applied to the whole plant decreased the number of trichomes and increased the BCP concentration from 0.79 to 1.7 mg g^{-1} in Mexican oregano (*Lippia graveolens*; Verbenaceae) (Eller, de Gouw, Graus, & Monson, 2012).

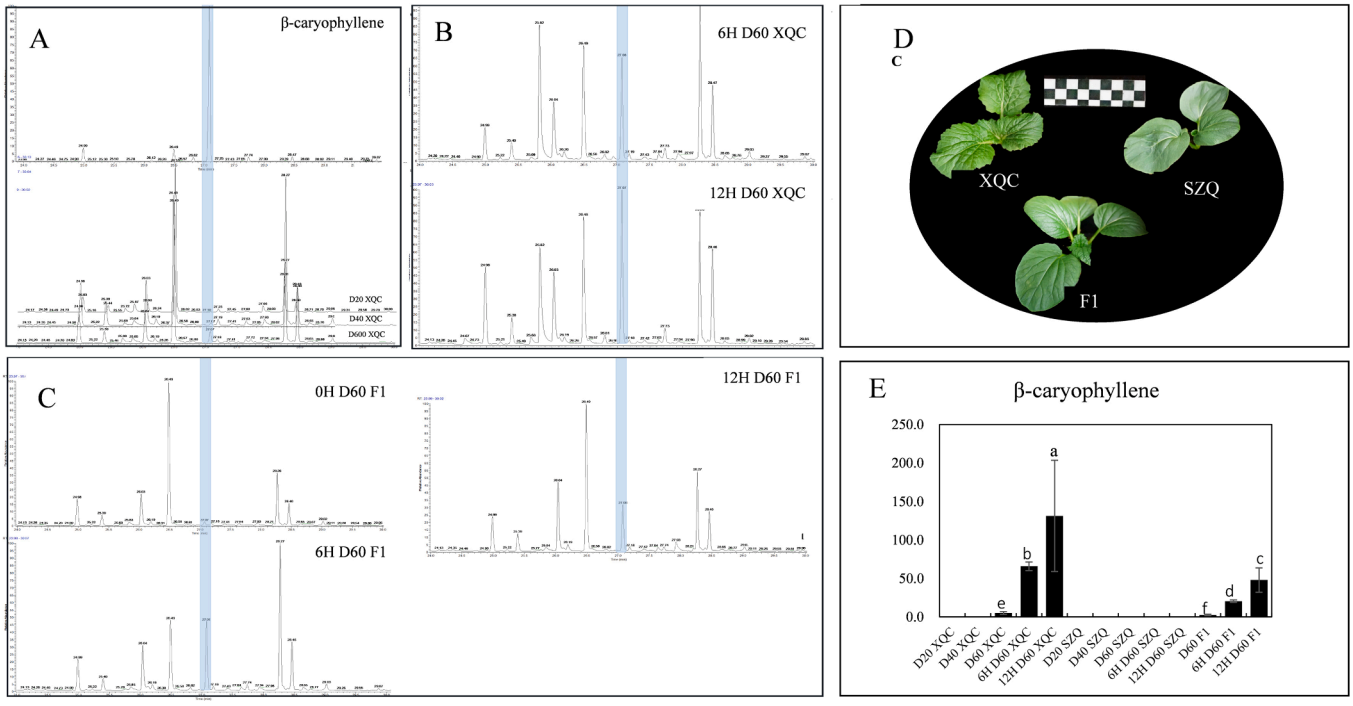


Fig. 2. GC–MS analysis of products in *Brassica campestris* leaves. (A) GC–MS analysis of VOCs in 'XQC' leaves at 20, 40, and 60 days. (B) GC–MS analysis of VOCs in 'XQC' leaves after MeJA treatment. (C) GC–MS analysis of VOCs in F1 hybrids leaves after MeJA treatment. (D) Types and characteristics of the three study materials. (E) BCP content of different samples.

3.3. Phylogenetic classification and TPSa gene transcriptome

The TPS gene family comprises important structural genes for synthesizing plant terpenoids that have been identified in various plants, including carrot (*D. carota*) (Keilwagen et al., 2017), grape (*V. vinifera*) (Schwab & Wust, 2015), apple (Nieuwenhuizen et al., 2013), cotton (Huang et al., 2013), and agarwood (*Aquilaria agallochum*) (Das et al., 2021). However, the TPS gene family has not been engineered in *B. campestris*. To examine the evolutionary relationships among the TPS genes of the *B. campestris*, *Arabidopsis*, and other species, the various TPS gene sequences were isolated from *B. campestris* (present study) (33), *Raphanus sativus* (52), *V. vinifera* (66), *B. napus* (60), *B. oleracea* (33), *B. rapa* (37), *Camelina sativa* (78), and *Arabidopsis* (32) (https://plants.ensembl.org/index.html) (Table S1).

By constructing the phylogenetic tree, we obtained 21 TPSa genes in NHCC001 and 201 TPSa genes in seven other plant species (Fig. 4A). To further understand the phylogenetic relationships of NHCC TPSa family genes, the full-length TPSa protein sequences of the eight species and 20 BCP synthase proteins in the NCBI database (as determined by a BLAST search) were used to construct the phylogenetic tree (Table S2). A total of 19 *BcTPSa* genes could be mapped on the linkage groups and were renamed from *BcTPSa1* to *BcTPSa19* based on their order in the linkage groups. Moreover, two TPS genes (*BraCxxg001060.1* and *BraCxxg006940.1*) that could not be conclusively mapped to NHCC were renamed *BcTPSa20* and *BcTPSa21*, respectively **(Fig. S2)**. Basic information on these genes is provided in Table S3. Comparative genomic analysis confirmed that NHCC001 underwent a whole-genome triplication (WGT) event shared with other *Brassica* species, which occurred after the whole-genome duplication (WGD) event shared with *Arabidopsis* (Li et al., 2020). Many genes of the TPSa gene subfamily members were lost after the WGD event, which was why the number of their members did not triple (Table S4). However, the BCP synthase gene was well conserved, and NHCC001 had a special BCP synthase gene homolog BraC03g066260.1, which was not identified as TPS because of only one domain. The exon–intron structure is an important evolutionary feature of a gene and holds clues to its functional diversification. Thus, the gene structure of these TPSa genes was further analyzed **(**Fig. 3**B)**. The 21 TPSa gene family members of NHCC001 possessed intron sequences.

To investigate the expression dynamics of the terpene biosynthesis genes, TPSa expression was evaluated. The expression patterns of all 21 TPSa genes in the transcriptome data, which was derived from different NCHH materials, were investigated in this study. Among the 21 TPSa genes, many TPSa genes were not expressed in all samples, suggesting that these genes may have spatiotemporal expression patterns that were not examined in this study. Interestingly, two genes, *BcTPSa11* and *BcTPSa21*, were expressed in all 'XQC' and F1 samples (TPM $>$ 1) and showed constitutive expression. However, transcription of these genes was not detected in 'SZQ' **(**Fig. 3**C)**. These evidences suggest that the *BcTPSa11* and *BcTPSa21* genes encoded key enzymes in BCP synthesis in 'XQC'. Furthermore, *BcTPSa21* encoded a rate-limiting enzyme in the BCP synthetic pathway in *B. campestris*, and transcriptome data was verified by qRT-PCR analysis **(Fig. S7, Table S11)**.

3.4. Phylogenetic analysis of BCP

To further understand the phylogenetic relationships of BCP genes, multiple sequence alignment of 29 BCP gene members from *Zea mays*, *O. sativa japonic*, *A. thalian*, *V. vinifera*, *Zanthoxylum ailanthoides*, *Citrus bergamia*, *S. lycopersicum*, *Piper nigrum*, *B. oleracea*, *B. rapa* 'Chiifu', *B. rapa* 'Z1′, and *B. rapa* NHCC001 was performed using OrthoFinder, and a maximum likelihood phylogenetic tree was generated based on the alignment result (Fig. 4B and Table S5). TPSa genes were classified into four groups (TPSa I–IV) (Fig. 4A). At least five TPSa genes from NHCC001 were present in each group, with the exception of TPSa genes in group TPSa II (Fig. 3B). Interestingly, BCP genes of five species were clustered, suggesting that differences in BCP members occurred after the

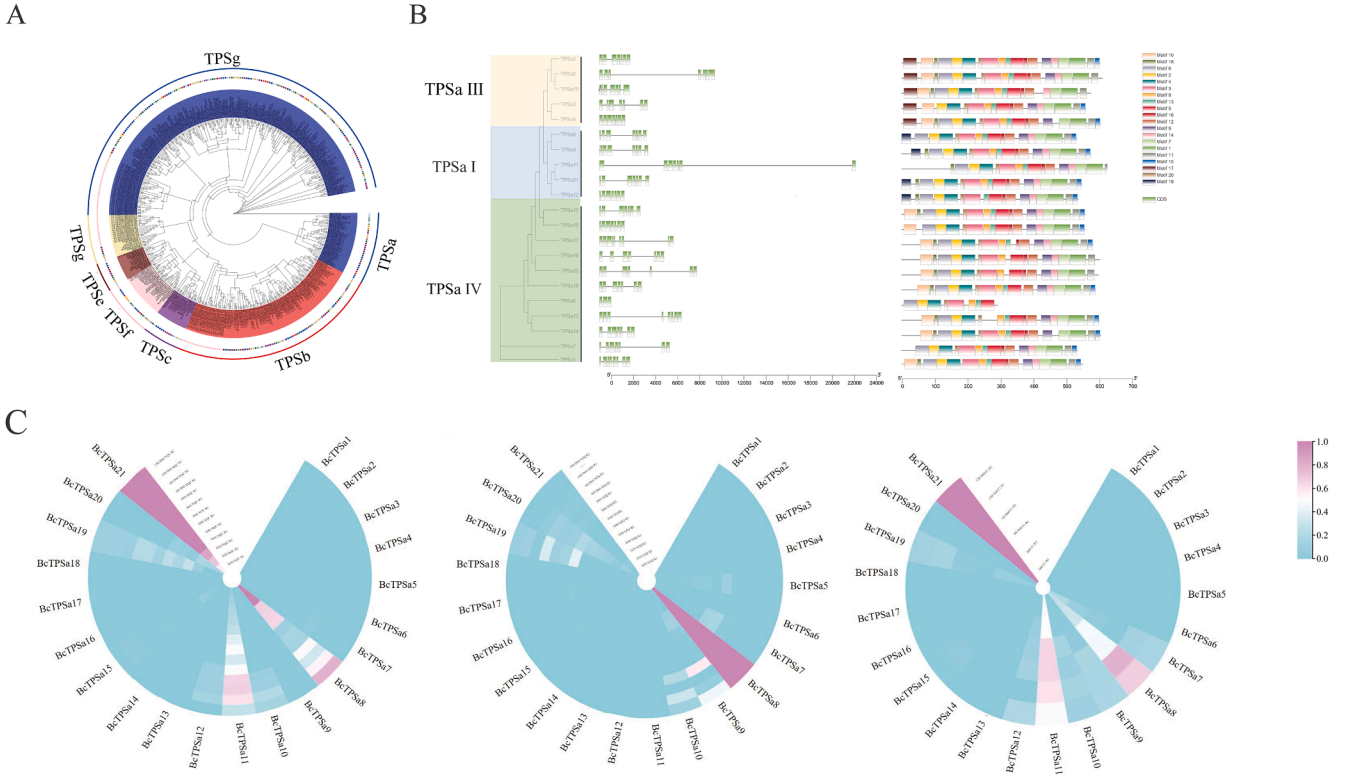


Fig. 3. Phylogenetic relationships, gene structure, and classification of *BcTPSa* genes. (A) Phylogeny of putative full-length TPS genes from eight sequenced plant genomes. (B) Phylogenetic relationships and gene structure in TPSa genes from *Brassica campestris.* (C) Expression of terpene biosynthesis genes in 'XQC' (left), 'SZQ' (middle), and F1 hybrids (right).

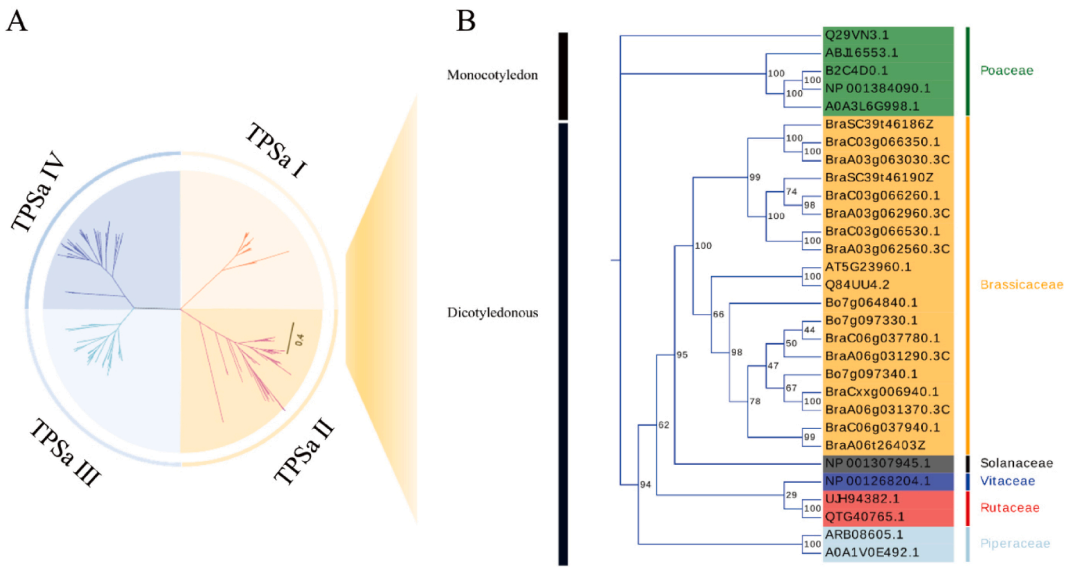


Fig. 4. Phylogenetic relationships of BCP proteins. (A) Four groups of TPSa genes. (B) Phylogenetic tree representing relationships among BCP genes from twelve species.

divergence of Brassicaceae and Solanaceae. A WGT event at 4Dtv ≈ 0.15 (*Ks* peak value ≈ 0.3), which was previously reported as a Brassicaceae-specific triplication (Br-α-WGD), had also occurred in the evolutionary history of NHCC001, and there was almost complete triplication of the NHCC001 genome relative to *A. thaliana* (Li et al., 2020). BCP members of the *Brassica* and *Arabidopsis* phylogenetic trees had six genes each, and dispersion might be an important reason for BCP gene expansion in members of Brassicaceae, which also implies the importance of these genes in BCP synthesis. Furthermore, BraC03g06626.1, an AT5g29360.1 ortholog determined by OrthoFinder, was discarded because of only one TPS domain.

3.5. Terpene synthesis pathway genes and candidate BCP transporters

To explore the transcription of terpene synthesis pathway genes, genes in both 'XQC' and 'SZQ' were evaluated for differential expression at different growth stages. Moreover, genes in parents and F1 were evaluated after MeJA treatment. The expression levels of each gene were calculated and normalized to TPM values. Multiple testing correction was performed with a Padjust value < 0.05, and 3,972 genes in 'XQC', 4,339 genes in 'SZQ', and 852 genes in F1 exhibited differential expression. The results of the Venn diagram and KEGG enrichment analysis of these DEGs are shown in Figure S3.

Fig. 5 shows the correlation heatmap of DEGs involved in the terpene biosynthetic pathway. In this study, 16 unigenes that encode enzymes associated with terpenes biosynthesis were identified based on the enriched KEGG pathways and GO functional analysis. These genes included Acetyl-CoA C-acetyltransferase *AACT* (1), HMG-CoA synthase *HMGC* (1), HMG-CoA reductase *HMGR* (2), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase *MDS* (1), Isopentenyl diphosphate isomerase *IDI* (2), 1-deoxy-D-xylulose-5-phosphate synthase *DXS* (1), 1-deoxy-D-xylulose-5-phosphate DXS reductoisomerase *DXR* (2), 4-hydroxy-3-methylbut-2-enyl diphosphate synthase *HDS* (1), and Terpenes synthase *TPS* (5) (Fig. 5A). Three TPS unigenes (*BraCxxg006940.1*, *BraC06g037780.1*, and *BraC03g069360.1*), one HMGR unigene (*BraC02g024430.1*), one DXS unigene (*BraC06g032300.1*), and one HDS unigene (*BraC10g019760.1*) were significantly correlated with BCP, respectively (Fig. 5B).

Transporters regulate the storage and transport of sesquiterpenes. In the trend analysis, six gene sets from F1 and four gene sets from 'XQC' were present in modules 18 and 19, respectively (Fig. S4). F1 and 'XQC' had 313 and 1,398 trend genes, respectively (Table S6). Correlation analysis was performed on all DEGs, and genes with a correlation greater than 0.9 were selected for KEGG and GO enrichment analyses (Fig. S5, Table S7, 8). DEGs enriched in transporters were annotated, and 18 transporters were screened, including those from the MFS family (5), OPT family (2), POT/PTR family (2), SulP family (2), ABC superfamily (1), MIP (3), MOP superfamily (1), MC (1), and ArAE family (1) (Table S9).

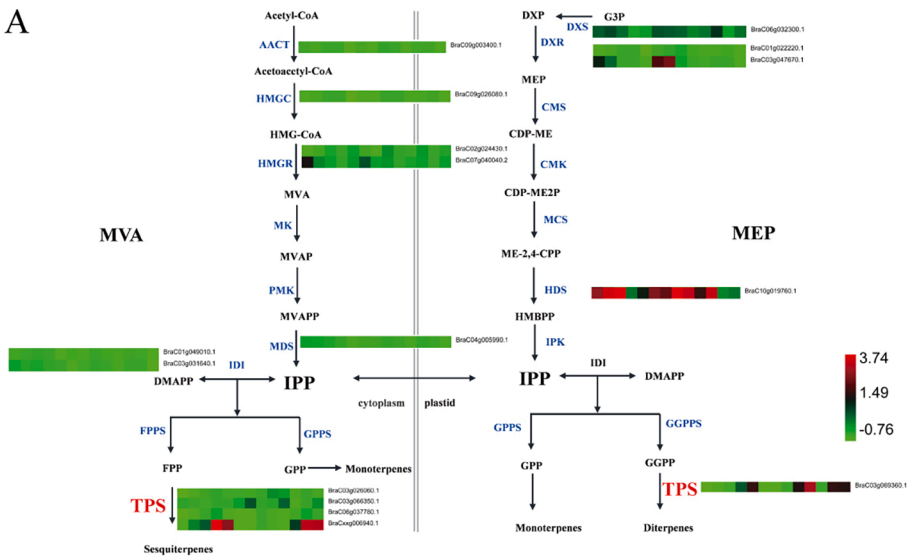
3.6. Identification of candidate TFs and subcellular localization and promoter activity analysis

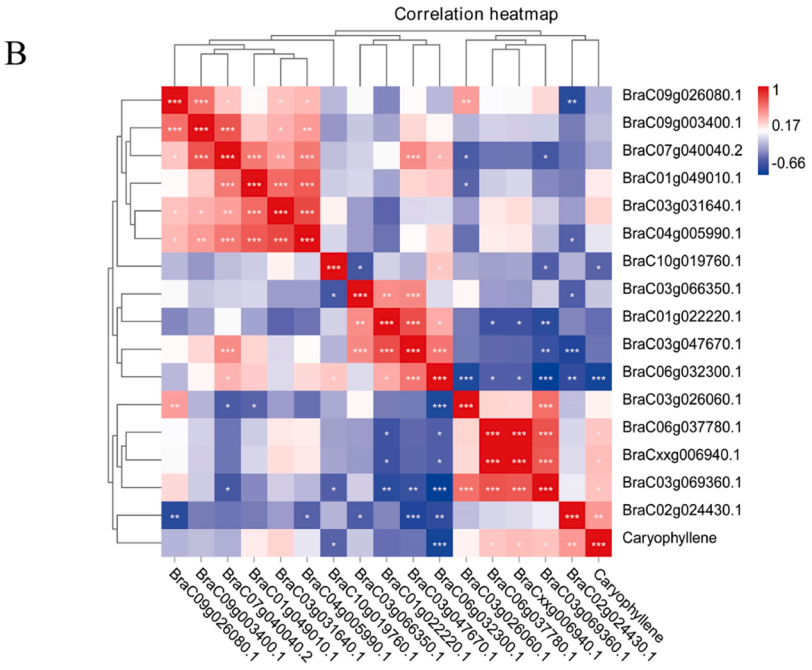
Differential expression profiles of TFs were analyzed. TFs have been identified in regulating the biosynthesis of terpene-specialized metabolites in different plants, and include AaWRKY, AaERF1, AaMYB1, AabZIP1 and MsYABBY5 (Shu et al., 2022; Q. Wang et al., 2016). DEGs related to TFs were 51 AP2/ERF, 33 bHLH, 15 MYB, 7 NF-Y, 7 TCP, 6 HSF EREBP, 22 WRKY, 5 MYB-related, 4 bZIP, 4 NAC, 3 trihelix, 1 GATA, 1 GRAS, 1 PCL, and 1 VOZ (Fig. S6, Table S10).

TFs are critical in regulating gene expression. To determine TFs that were co-expressed with the 16 unigenes, Pearson's correlation analysis was performed between the expression levels of all DEGs annotated as TFs and candidate genes. TFs with Pearson's correlation coefficients ≥ 0.9 were selected (Fig. 6, Table S12). TFs related to *BcTPSa21* gene were determined by subcellular localization and a transient dual-luciferase (dual-LUC) experiment in tobacco leaves. Green fluorescent protein-fused BcERF23, BcMYB122, and BcMYC2 reporter genes were constructed and used for transient transformation experiments in tobacco cells, and they were found to be nuclear localization proteins (Fig. 6**C**). The dual-LUC experiment showed that BcERF23, BcMYB122, and BcMYC2 might bind the promoter of TPSa21–300 (Fig. 6**D**).

4. Discussion

β-Caryophyllene (BCP) is a bicyclic sesquiterpene widely distributed in the plant kingdom. It provides a unique aroma to essential oils and plays a pivotal role in the survival and evolution of higher plants. This plant volatile compound is commonly found in grapes (*V. vinifera*) (Schwab & Wust, 2015), rose myrtle (*Rhodomyrtus tomentosa*) (He et al., 2018), black pepper (*P. nigrum*) (Jin et al., 2018), cannabis (*Cannabis sativa*) (Booth, Page, & Bohlmann, 2017), and Japanese chrysanthemum (*Chrysanthemum japonense*) (Usami, Ono, Marumoto, & Miyazawa, 2013). However, few studies have been conducted on aroma





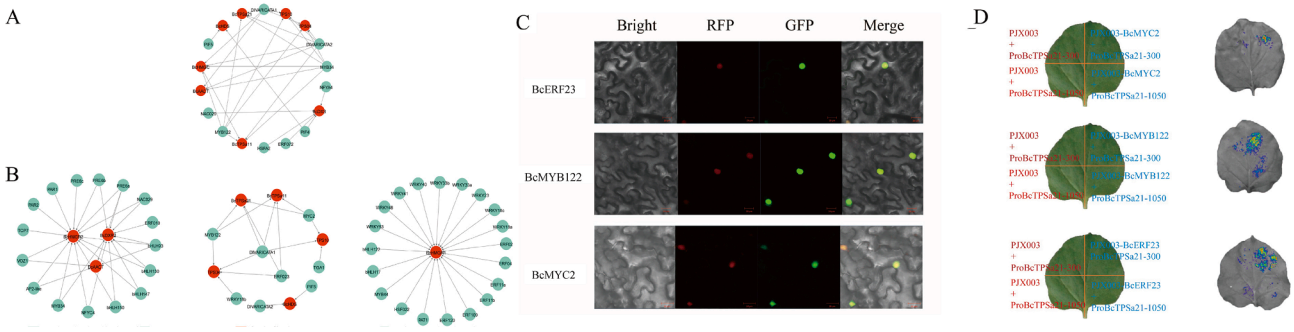


Fig. 5. Analysis of terpene biosynthetic pathway genes. (A) Terpene synthases characterized in this study are shown in bold. At, *Arabidopsis thaliana*; CPP, *ent*-copalyl diphosphate; DMAPP, dimethylallyl diphosphate; FPP, farnesyl diphosphate; GFPP, geranylfarnesyl diphosphate; GGPP, geranylgeranyl diphosphate; GPP, geranyl diphosphate; TPS, terpene synthase. (B) Heatmap generated using correlation coefficients between TPM values of terpene biosynthetic pathway genes and relative BCP content listed in Table S5. *, $p < 0.05$; **, $p < 0.01$.

Fig. 6. Putative regulation of TPSa. (A) Putative regulation in F1 hybrids. (B) Putative regulation in 'XQC'. (C) Nuclear localization of BcERF23, BcMYB122, and BcMYC2 in tobacco leaf cells. (D) Analysis of the promoter activity of BcERF23, BcMYB122, and BcMYC2 in tobacco leaves.

compounds, such as BCP, in *B. campestris*. In this study, we identified the sesquiterpene BCP in *B. campestris* by sampling headspace volatiles with a homemade sampler.

Since most studies on the genetic control of flavor-related traits focus on genotypes, the *B. campestris* cultivars 'XQC' and 'SZQ' were crossed to produce F1 hybrids. 65 % of VOC content strongly correlated between parental lines and hybrids (Bineau et al., 2021). Genotype and parental crosses were also significant predictors of isoprene and monoterpene emissions (Eller et al., 2012). Our results indicated that the genotype might be why the flavor-related trait of BCP was inherited to F1 hybrids **(**Fig. 2**)**.

BCPs of five species were clustered and differences in BCP members might occur between Brassicaceae and Solanaceae **(**Fig. 4**)**. A WGT event at 4Dtv ≈ 0.15 (K*s* peak value ≈ 0.3), which was previously reported as a Brassicaceae-specific triplication (Br-α-WGD), had also occurred in the evolutionary history of NHCC001, and the *B. rapa* NHCC001 genome had almost complete triplication relative to *A. thaliana* (Li et al., 2020). The results showed that BCP members of the *Brassica* phylogenetic tree had six genes each, and dispersion might be an important reason for the expansion of BCP in members of Brassicaceae **(**Fig. 4**B)**. Furthermore, BraC03g066260_1, an AT5g29360.1 ortholog found by OrthoFinder, was discarded due to only one TPS domain.

Geranyllinalool synthases can be induced in leaves with MeJA in Solanaceae and other angiosperms (Falara, Alba, Kant, Schuurink, & Pichersky, 2014). A rate-limiting step in MVA pathway involves 3-hydroxy-3-methylglutaryl-CoA reductase, which produces mevalonate (Chappell, Wolf, Proulx, Cuellar, & Saunders, 1995). We found that the BraC02g024430.1 gene, which synthesizes (E, E)-geranyllinalool, was significantly upregulated on MeJA induction.

BraC09g053570.1 (ABCG18) is a transporter related to TPSa21 **(Table S9)**, and its family member ABCG31 is related to AaPDR3, a pleiotropic drug resistance (PDR) transporter 3, involved in sesquiterpene BCP transport in *Artemisia annua* (Fu et al., 2017). Moreover, ABCG29 is involved in diterpene transport to defend against biotic threats (Pierman et al., 2017). ABC transporters are involved in the transport of secondary metabolites in medicinal plants (Liu, Sun, Liu, Shao, & Zhang, 2021). A total of 49 ABC transporter genes have been divided into eight subfamilies (ABCA–ABCH), including seven *ABCA* genes, seven *ABCB* genes, 10 *ABCC* genes, two *ABCD* genes, one *ABCE* gene, three *ABCF* genes, 16 *ABCG* genes, and three *ABCH* genes according to phylogenetic analysis in melon fly (*Zeugodacus cucurbitae*), a highly destructive insect pest of cucurbitaceous and other related crops (Xu et al., 2021). Thus, the PDR family ABCG might be involved in BCP transport in NHCC.

From the selected TFs, MYC2 TFs closely correlated with the candidate unigenes *BcTPSa21* and *BcTPSa11*. *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression (Hong et al., 2012). In addition to MYC2 TFs, DIVARICATA1, DIVARICATA2, MYB122, ERF023, and MYB34 TFs closely correlated with the *BcTPSa21* and *BcTPSa11* genes, suggesting that these TFs are candidate regulators for sesquiterpene biosynthesis **(**Fig. 6**)**. Several TFs regulate sesquiterpene biosynthesis in plants. MrTPS3 and MrTPS20 are responsible for BCP and α-pinene production, respectively, in red bayberry (Wang et al., 2022). *CpbHLH13* overexpression in *A. thaliana* ecotype Columbia-0 and tobacco genotype SR1 increased floral sesquiterpene BCP production in both transgenic plants (Aslam, Lin, Li, Yang, & Chen, 2020). Downregulation of SlMYB75 increased the formation of type II, V, and VI trichomes and the accumulation of δ-elemene, BCP, and α-humulene in glandular trichomes (Gong et al., 2021). The miR156-targeted SPL transcription factor plays an important role in the spatiotemporal regulation of sesquiterpene biosynthesis. In *A. thaliana*, SPL9 directly binds to the TPS21 promoter and activates its expression (Yu et al., 2015). In this study, Pearson correlation analysis between transcripts annotated as TFs and candidate unigenes revealed that the expression levels of some TFs closely correlated with terpene biosynthetic pathway genes, indicating that these TFs might regulate the expression of terpene biosynthetic pathway genes and indirectly contribute to terpene accumulation in *B. campestris*.

5. Conclusion

Volatile organic compounds are an important characteristic of vegetable flavor and maturity. In this study, VOCs provided information on *B. campestris* aroma composition and aroma compound formation, which can guide *B. campestris* plant breeding. In addition, the composition and the regulatory mechanism of the synthesis of aroma compounds released by *B. campestris* were revealed by using the combination of HS-SPME/GC–MS and RNA-seq. The changes in accumulation levels and gene expression levels at 60 days elucidated aroma compound formation during vegetable ripening. BCP, which can be induced by MeJA, is the aroma component released by 'XQC', and the trait is transmitted to hybrids. Gene expression at different developmental stages also supports our theory.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Ph.D. Bai Yibo from Nanjing Agricultural University for their constructive comments and Shen sheng yuan who provide material for this study. This work was supported by the National Key Research and Development Program (2018YFD1000805), Jiangsu Seed Industry Revitalization Project [JBGS (2021)015], National vegetable industry technology system (CARS-23-A-16), and a project funded by the priority academic program development of Jiangsu higher education institutions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2022.100129.

References

Aslam, M. Z., Lin, X., Li, X., Yang, N., & Chen, L. (2020). Molecular cloning and functional characterization of CpMYC2 and CpBHLH13 transcription factors from wintersweet (*Chimonanthus praeco*x L.). *Plants, 9*(6), 785.

Aubourg, S., Lecharny, A., & Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Molecular Genetics and Genomics, 267*(6), 730–745.

Bineau, E., Rambla, J. L., Priego-Cubero, S., Hereil, A., Bitton, F., Plissonneau, C., … Causse, M. (2021). Breeding Tomato Hybrids for Flavour: Comparison of GWAS results obtained on lines and F1 hybrids. *Genes, 12*(9), 1443.

Bohlmann, J., Meyer-Gauen, G., & Croteau, R. (1998). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proceedings of the National Academy of Sciences of the United States of America, 95*(8), 4126–4133.

Booth, J. K., Page, J. E., & Bohlmann, J. (2017). Terpene synthases from *Cannabis sativa*. *PloS one, 12*(3), Article e0173911.

Chappell, J., Wolf, F., Proulx, J., Cuellar, R., & Saunders, C. (1995). Is the reaction catalyzed by 3Hydroxy-3-Methylglutaryl coenzyme a reductase a aate-limiting step for isoprenoid biosynthesis in plants? *Plant Physiology, 109*(4), 1337–1343.

Chen, H., Li, G., Köllner, T. G., Jia, Q., Gershenzon, J., & Chen, F. (2014). Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus Oryza. *BMC Plant Biology, 14*, 239.

Chen, J., Zhang, J., Xiang, Y., Xiang, L., Liu, Y., He, X., … Huang, Z. (2016). Extracts of Tsai Tai (*Brassica chinensis*): Enhanced antioxidant activity and anti-aging effects both in vitro and in *Caenorhabditis elegans*. *Food & Function, 7*(2), 943–952.

Das, A., Begum, K., Akhtar, S., Ahmed, R., Kulkarni, R., & Banu, S. (2021). Genome-wide detection and classification of terpene synthase genes in *Aquilaria agallochum*. *Physiology and Molecular Biology of Plants, 27*(8), 1711–1729.

Eller, A. S., de Gouw, J., Graus, M., & Monson, R. K. (2012). Variation among different genotypes of hybrid poplar with regard to leaf volatile organic compound emissions. *Ecological Applications, 22*(7), 18651875.

Falara, V., Akhtar, T. A., Nguyen, T. T., Spyropoulou, E. A., Bleeker, P. M., Schauvinhold, I., … Pichersky, E. (2011). The tomato terpene synthase gene family. *Plant physiology, 157*(2), 770–789.

Falara, V., Alba, J. M., Kant, M. R., Schuurink, R. C., & Pichersky, E. (2014). Geranyllinalool synthases in solanaceae and other angiosperms constitute an ancient branch of diterpene synthases involved in the synthesis of defensive compounds. *Plant Physiology, 166*(1), 428–441.

Fu, X., Shi, P., He, Q., Shen, Q., Tang, Y., Pan, Q., … Tang, K. (2017). AaPDR3, a PDR transporter 3, is involved in sesquiterpene β-caryophyllene transport in Artemisia annua. *Frontiers in Plant Science, 8*, 723.

Gong, Z., Luo, Y., Zhang, W., Jian, W., Zhang, L., Gao, X., … Deng, W. (2021). A SlMYB75-centred transcriptional cascade regulates trichome formation and sesquiterpene accumulation in tomato. *Journal of Experimental Botany, 72*(10), 3806–3820.

He, S. M., Wang, X., Yang, S. C., Dong, Y., Zhao, Q. M., Yang, J. L., … Fan, W. (2018). *De novo* transcriptome characterization of *Rhodomyrtus tomentosa* leaves and identification of genes involved in α/β-pinene and β-caryophyllene biosynthesis. *Frontiers in Plant Science, 9*, 1231.

Hong, G. J., Xue, X. Y., Mao, Y. B., Wang, L. J., & Chen, X. Y. (2012). Arabidopsis MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *The Plant cell, 24*(6), 2635–2648.

Huang, X., Xiao, Y., Köllner, T. G., Zhang, W., Wu, J., Wu, J., … Zhang, Y. (2013). Identification and characterization of (E)-β-caryophyllene synthase and α/β-pinene synthase potentially involved in constitutive and herbivore-induced terpene formation in cotton. *Plant Physiology and Biochemistry, 73*, 302–308.

Irmisch, S., Jiang, Y., Chen, F., Gershenzon, J., & Köllner, T. G. (2014). Terpene synthases and their contribution to herbivore-induced volatile emission in western balsam poplar (*Populus trichocarpa*). *BMC Plant Biology, 14*, 270.

Jin, Z., Kwon, M., Lee, A. R., Ro, D. K., Wungsintaweekul, J., & Kim, S. U. (2018). Molecular cloning and functional characterization of three terpene synthases from unripe fruit of black pepper (*Piper nigrum*). *Archives of biochemistry and biophysics, 638*, 35–40.

Keilwagen, J., Lehnert, H., Berner, T., Budahn, H., Nothnagel, T., Ulrich, D., & Dunemann, F. (2017). The Terpene Synthase Gene Family of Carrot (Daucus carota L.): Identification of QTLs and candidate genes associated with terpenoid volatile compounds. *Frontiers in Plant Science, 8*, 1930.

Li, Y., Liu, G. F., Ma, L. M., Liu, T. K., Zhang, C. W., Xiao, D., … Hou, X. L. (2020). A chromosome-level reference genome of non-heading Chinese cabbage [*Brassica campestris* (syn. *Brassica rapa*) ssp. *chinensis*]. *Horticulture Research, 7*(1), 212.

Liu, J., Huang, F., Wang, X., Zhang, M., Zheng, R., Wang, J., & Yu, D. (2014). Genome-wide analysis of terpene synthases in soybean: Functional characterization of GmTPS3. *Gene, 544*(1), 83–92.

Liu, N., Sun, H., Liu, Z. B., Shao, C., & Zhang, L. L. (2021). Progress of studies on ATP-binding cassette transporters and transportation of secondary metabolites in medicinal plants. *Zhongguo Zhong Yao Za Zhi, 46*(1), 52–56. in Chinese.

Martin, D. M., Aubourg, S., Schouwey, M. B., Daviet, L., Schalk, M., Toub, O., … Bohlmann, J. (2010). Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biology, 10*, 226.

Nieuwenhuizen, N. J., Green, S. A., Chen, X., Bailleul, E. J., Matich, A. J., Wang, M. Y., & Atkinson, R. G. (2013). Functional genomics reveals that a compact terpene synthase gene family can account for terpene volatile production in apple. *Plant Physiology, 161*(2), 787–804.

Pierman, B., Toussaint, F., Bertin, A., Levy, D., Smargiasso, N., De Pauw, E., & Boutry, M. (2017). Activity of the purified plant ABC transporter NtPDR1 is stimulated by diterpenes and sesquiterpenes involved in constitutive and induced defenses. *Journal of Biological Chemistry, 292*(47), 1949119502.

Qu, Z., Zhang, L., Zhu, S., Yuan, W., Hang, J., Yin, D., … Sun, J. (2020). Overexpression of the transcription factor HAC1 improves nerolidol production in engineered yeast. *Enzyme and Microbial Technology, 134*, Article 109485.

Schwab, W., & Wust, M. (2015). Understanding the constitutive and induced biosynthesis of mono- and sesquiterpenes in grapes (*Vitis vinifera*): A key to unlocking the biochemical secrets of unique grape aroma profiles. *Journal of Agricultural and Food Chemistry, 63*(49), 10591–10603.

Sharon-Asa, L., Shalit, M., Frydman, A., Bar, E., Holland, D., Or, E., … Eyal, Y. (2003). Citrus fruit flavor and aroma biosynthesis: Isolation, functional characterization, and developmental regulation of *Cstps1*, a key gene in the production of the sesquiterpene aroma compound valencene. *The Plant Journal, 36*(5), 664–674.

Shu, G., Tang, Y., Yuan, M., Wei, N., Zhang, F., Yang, C., … Liao, Z. (2022). Molecular insights into AabZIP1-mediated regulation on artemisinin biosynthesis and drought tolerance in *Artemisia annua*. *Acta Pharmaceutica Sinica. B, 12*(3), 1500–1513.

Usami, A., Ono, T., Marumoto, S., & Miyazawa, M. (2013). Comparison of volatile compounds with characteristic odor in flowers and leaves of nojigiku (*Chrysanthemum japonense*). *Journal of Oleo Science, 62*(8), 631–636.

Wang, Q., Reddy, V. A., Panicker, D., Mao, H. Z., Kumar, N., Rajan, C., … Sarojam, R. (2016). Metabolic engineering of terpene biosynthesis in plants using a trichomespecific transcription factor MsYABBY5 from spearmint (*Mentha spicata*). *Plant Biotechnology Journal, 14*(7), 1619–1632.

Wang, Y., Yang, Q., Zhu, Y., Zhao, L., Ju, P., Wang, G., … Gao, Z. (2022). MrTPS3 and MrTPS20 are responsible for β-caryophyllene and α-pinene production, respectively, in red bayberry (*Morella rubra*). *Frontiers in Plant Science, 12*, Article 798086.

Xu, H. Q., Ma, M., Ma, Y. P., Zhang, S. Y., Li, W. J., Wei, D., & Wang, J. J. (2021). Identification and expression characterization of ATP-binding cassette (ABC) transporter genes in melon fly. *Insects, 12*(3).

Yang, Z., Li, Y., Gao, F., Jin, W., Li, S., Kimani, S., … Wang, L. (2020). MYB21 interacts with MYC2 to control the expression of terpene synthase genes in flowers of *Freesia hybrida* and Arabidopsis *thaliana*. *Journal of Experimental Botany, 71*(14), 4140–4158.

Yu, N., Chen, Z., Yang, J., Li, R., & Zou, W. (2021). Integrated transcriptomic and metabolomic analyses reveal regulation of terpene biosynthesis in the stems of *Sindora glabra*. *Tree Physiology, 41*(6), 1087–1102.

Yu, Z. X., Wang, L. J., Zhao, B., Shan, C. M., Zhang, Y. H., Chen, D. F., & Chen, X. Y. (2015). Progressive regulation of sesquiterpene biosynthesis in *Arabidopsis* and patchouli (*Pogostemon cablin*) by the miR156-targeted SPL transcription factors. *Molecular Plant, 8*(1), 98–110.

Yu, N., Yang, J. C., Yin, G. T., Li, R. S., & Zou, W. T. (2018). Transcriptome analysis of oleoresinproducing tree *Sindora glabra* and characterization of sesquiterpene synthases. *Frontiers in Plant Science, 9*, 1619.