Article

Integration of Metabolite Profiling and Transcriptome Analysis Reveals Genes Related to Volatile Terpenoid Metabolism in Finger Citron (*C. medica* var. *sarcodactylis*)

Yaying Xu [1], Changqing Zhu [1], Changjie Xu [1], Jun Sun [2], Donald Grierson [1,3], Bo Zhang [1,*] and Kunsong Chen [1]

[1] Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology/Laboratory of Fruit Quality Biology, Zhejiang University, Hangzhou 310058, China

[2] Zhejiang Agricultural Technology Extension Center, Hangzhou 310029, China

[3] Plant and Crop Sciences Division, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire LE12 5RD, UK

* Correspondence: bozhang@zju.edu.cn; Tel.: +86-571-88982471; Fax: 86-571-88982224

check for updates

Abstract: Finger citron (*Citrus medica* var. *sarcodactylis*) is a popular ornamental tree and an important source of essential oils rich in terpenoids, but the mechanisms behind volatile formation are poorly understood. We investigated gene expression changes combined with volatile profiling of ten samples from three developing organs: flower, leaf, and fruit. A total of 62 volatiles were identified with limonene and γ-terpinene being the most abundant ones. Six volatiles were identified using partial least squares discriminant analysis (PLS-DA) that could be used as markers for distinguishing finger citron from other citrus species. RNA-Seq revealed 1,611,966,118 high quality clean reads that were assembled into 32,579 unigenes. From these a total of 58 terpene synthase (TPS) gene family members were identified and the spatial and temporal distribution of their transcripts was measured in developing organs. Transcript levels of transcription factor genes *AP2/ERF* (251), *bHLH* (169), *bZIP* (76), *MYB* (155), *NAC* (184), and *WRKY* (66) during finger citron development were also analyzed. From extracted subnetworks of three modules constructed by weighted gene co-expression network analysis (WGCNA), thirteen *TPS* genes and fifteen transcription factors were suggested to be related to volatile terpenoid formation. These results provide a framework for future investigations into the identification and regulatory network of terpenoids in finger citron.

Keywords: citrus; limonene; γ-terpinene; WGCNA; terpene synthase gene; transcription factor; RNA-Seq

1. Introduction

Plants produce a range of small molecule volatile and non-volatile organic compounds called secondary metabolites that are not essential to the growth and development of the producing organism. But enable them to adapt to their living environment. Terpenoids constitute the largest group of volatile secondary metabolites [1]. They are made from the basic C5 building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) and are synthesized by the mevalonate (MVA) pathway in the cytosol and the methyl-erythritol phosphate (MEP) pathway in the plastid [2]. IPP and DMAPP are condensed by head to tail reactions to form geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) [3]. Terpene synthases (TPS) function to catalyze the terminal steps in these pathways and utilize substrates GPP, FPP and GGPP to form monoterpene

(C10), sesquiterpene (C15), and diterpene (C20), respectively. The diversity of terpene synthase (TPS) gene members and their catalytic mechanisms generate extremely complex products [4]. Monoterpenes and sesquiterpenes are volatile at room temperature. They are important secondary metabolites with biological functions such as promoting stress resistance or acting as plant pheromones [5,6]. Moreover, monoterpenoids and sesquiterpenoids contribute to the characteristic aroma of fruits, including citrus fruits [7].

TPS families have been identified in many plants, for example Arabidopsis, tomato, grapevine, and apple [8–11]. Chen et al. comprehensively introduced the structure, function, and evolution of plant TPS [12]. The modern plant TPS family went through several evolutionary events such as gene duplication, functional differentiation, and divergence, and finally developed into seven subfamilies [8,13]. Dornelas et al. found 49 TPS genes from multiple citrus varieties by searching the CitEST database [14]. With the publication of the *Citrus sinensis* genome, Alquézar et al. described the first functionally characterized TPS family in citrus [15]. They identified seven genes which encode sesquiterpene synthases. Other terpene synthases from Citrus spp. have been functionally characterized in mandarin [16–20], lemon [21], sweet orange [22], and yuzu (*Citrus junos*) [23]. Recently, we have identified a sesquiterpene synthase associated with formation of bicyclogermacrene in finger citron [24].

Terpene biosynthesis changes during plant development. Progress has been made in identifying transcription factors (TFs) related to regulation of terpenoid synthesis. In citrus, CitAP2.10 [25] and CitERF71 [26] were reported to regulate the synthesis of the sesquiterpene valencene and monoterpenoid (*E*)-geraniol, respectively. MsMYB negatively regulates monoterpene synthesis in spearmint trichomes [27]. Nieuwenhuizen et al. found that mutation in the NAC binding region of TPS gene promoters in two kiwifruit species affect their monoterpene contents [28]. In cotton, GaWRKY1 up-regulates the expression of *CAD1*, leading to the accumulation of sesquiterpene phytoalexins, including gossypol [29] while overexpression of *OsbZIP79* causes a decrease in the accumulation of phytoalexin in rice cells [30]. Hong et al. found that AtMYC2 combines multiple plant hormone signals in the transcriptional regulation of sesquiterpene biosynthesis and activates the production of (*E*)-β-caryophyllene [31]. AaERF1 and AaERF2 positively regulate the expression level of *ADS* and *CYP71AV1*, leading to the accumulation of artemisinin [32,33].

Citron is one of the three primitive species of Citrus [34]. The carpels of finger citron split, causing a finger-like fruit shape [35]. It is a popular ornamental tree in oriental countries with dietary and pharmaceutical potentials [36,37]. Finger citron essential oils have antioxidant, antibacterial, antibiofilm activities [38–40] and they have been widely used in the food and perfume industry [41]. The chemicals isolated from finger citron essential oils are terpenes, terpene derivatives, and various volatile substances such as higher alcohols, aldehydes, ketones, and esters [38]. The most abundant components of finger citron essential oils are limonene and γ-terpinene [39,40]. Compared to other Citrus species, *C. medica* (citron) has a unique volatile profile [25,42,43] and according to González-Mas et al., it produces higher amounts of γ-terpinene, β-pinene or camphene, while lacking some compounds, especially non-terpenoid aldehydes [7].

Information on finger citron genes related to production of volatiles is still very limited, however, but now the availability of the *C. medica* genome provides essential reference for the investigation into gene function [34]. In the present study, we utilized metabolomic and transcriptomic methods to explore terpenoid biosynthesis and regulation in finger citron. Volatile profiles from developing finger citron organs were evaluated using GC-MS. Characteristic volatile compounds were identified by multivariate strategies. Also, the expression patterns of TPS, and TFs AP2/ERF, bHLH, MYB, NAC, and WRKY were described. We used weighted correlation network analysis (WGCNA) to uncover the correlations between terpene content and expression levels of specific genes [44]. This method has been successfully applied in mining potential TFs in grapevine and sugarcane [45,46] and elucidating the coordinated regulation of compounds in tea plants [47]. This strategy reconstructed subnetworks

with high connectivities, presenting gene targets for functional characterization. This information will contribute to development of the citrus industry and provide information for future breeding programs.

2. Results

2.1. Volatile Profiling of Finger Citron

In order to obtain comprehensive information about volatile components of finger citron, flowers, leaves and fruits at different developmental stages were subjected to SPME-GC-MS analysis (Figure 1A). Finger citron fruit weight increased from 68.33 g to 260.77 g during development (Figure 1B). A total of 62 volatiles were detected in finger citron throughout these developmental stages (Figure 1C and Table S1). Major volatile classes of finger citron included 17 monoterpene hydrocarbons, 23 oxygenated monoterpenes, 13 sesquiterpene hydrocarbons, one oxygenated sesquiterpene, and 8 others including aldehydes, alkanes, ketones, and olefins (Table S1). Among these 62 volatiles identified, 33 chemicals (53.2%) were detected in all three organs (flower, leaf, and fruit). A total of eight chemicals (12.9%) were only detected in flower. Three chemicals (4.8%) were specifically produced by fruit and no unique volatiles were detected in finger citron leaf (Figure 1C). Monoterpene was the most abundant class, accounting for 64.82% to 90.97% of the total content, followed by monoterpene derivative accounting for 6.84% to 32.47% (Figure 1D). Finger citron accumulated volatile compounds during organ development, with the highest content of volatiles detected in fruit, followed by leaf and flower. Among all the volatiles detected, the most abundant chemical was limonene (3.88 mg/g to 16.73 mg/g), followed by γ-terpinene (0.92 mg/g to 7.22 mg/g) (Table S1).
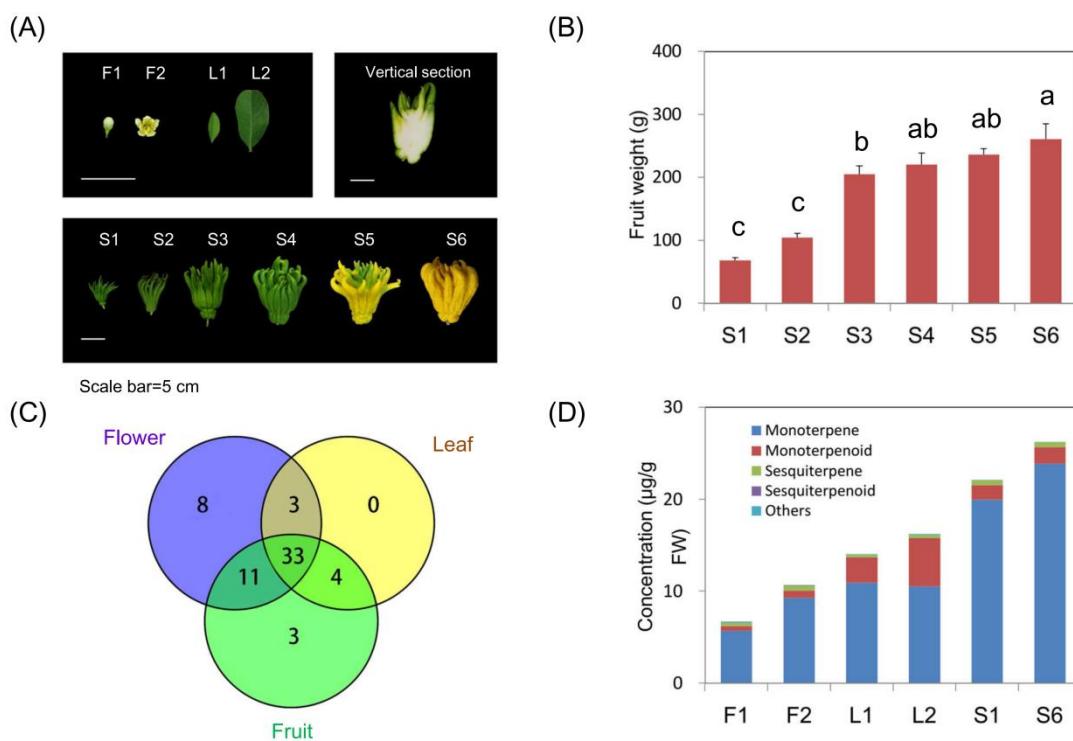

Figure 1. Major volatile classes from finger citron. (**A**) Photos of flower, leaf and fruit. (**B**) Changes in fruit weight during ripening. (**C**) Venn diagram of volatiles detected in three organs. (**D**) Concentration of major volatile classes in developing organs. F1, flower bud; F2, full flower, L1, young leaf; L2, mature leaf; S1–6, fruit developing stages 1–6. Scale bar = 5 cm. Significant differences between groups were detected by Duncan's test at significant level of 0.05.

A heat map was constructed to provide an overall view of volatiles detected in finger citron organs at different developmental stages (Figure 2). In brief, volatiles could be clustered into three

groups based on their profiles in finger citron organs. Group I volatiles were enriched in young leaf (F1). Volatiles in group II were enriched in fruit, including the two most abundant compounds, limonene and γ-terpinene. A total of 20 volatiles from group II were mainly detected in fruit and the contents of 17 volatiles accumulated during fruit development. In contrast, the content of the major volatile limonene tended to decrease during finger citron fruit ripening, declining from 16.73 mg/g to 13.41 mg/g (Figure 2 and Table S1). Group III consisted of 13 volatiles that mainly accumulated in fully opened flowers. The most notable change was observed for β-caryophyllene, whose contents increased more than 90-fold during flower maturation (Table S1).
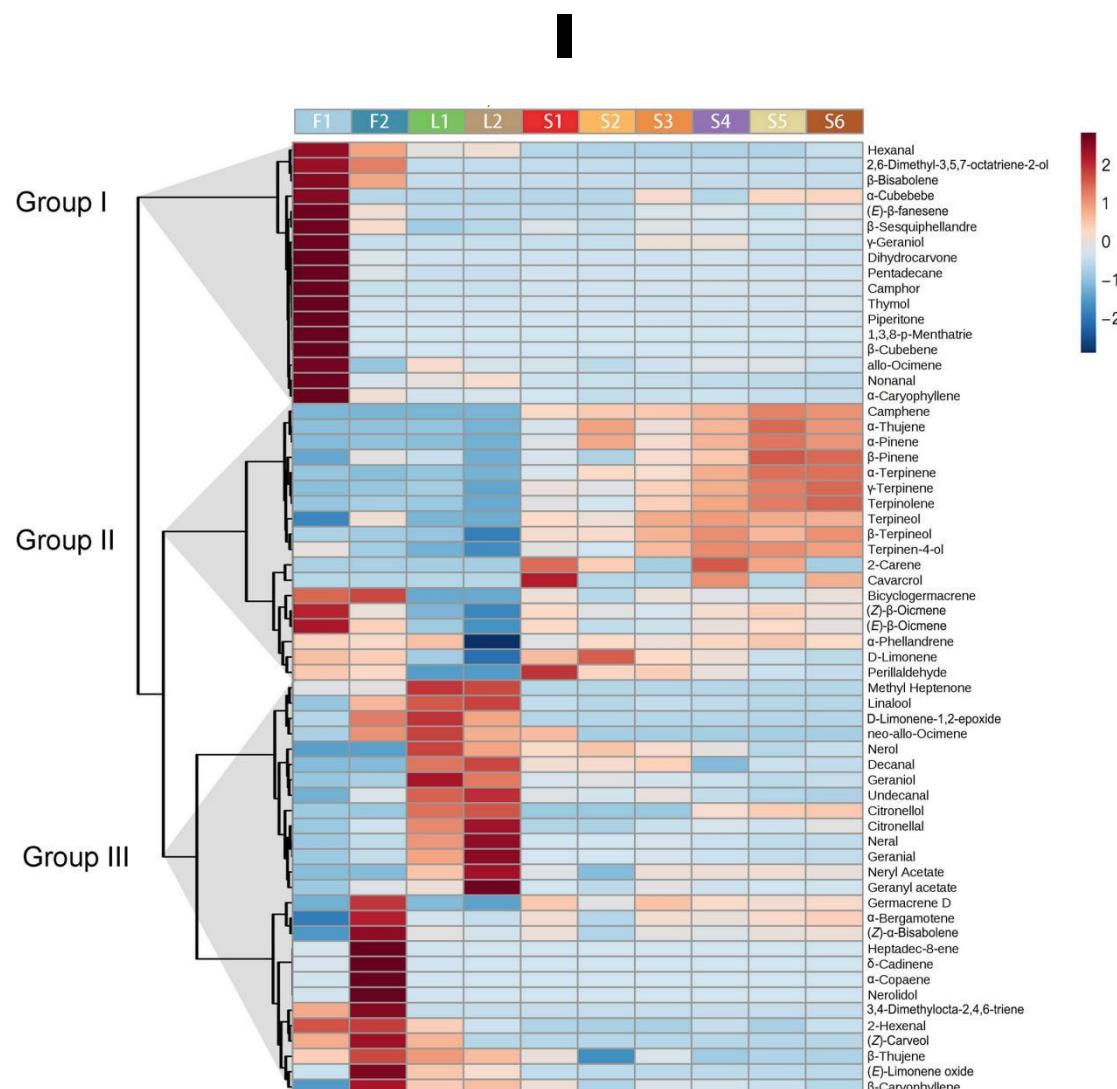

Figure 2. Hierarchical clustering and heat map visualization of changes in volatiles identified in developing finger citron. The left side of the heat map represents hierarchical clustering based on Pearson correlation. The color scale (−2 to 2) is shown on the right; red represents high content, blue represents low content. F1, flower bud; F2, full flower, L1, young leaf; L2, mature leaf; S1–6, fruit developing stages 1–6. Hierarchical clustering and heat map were constructed using MetaboAnalyst 4.0.

2.2. Identification of Key Volatiles

Variation in contents of volatiles observed in Figure 2 prompted us to investigate if finger citron organs could be distinguished based on volatile profiles. A PLS-DA model was constructed to explore the separation of organs using volatiles as variables. Permutation and cross validation tests validated the significance of the treatment differences with $p < 0.01$ (Figure S1A,B), indicating that the obtained

PLS-DA model is reliable. A separation of the finger citron leaf, flower, and fruit was observed in the PLS-DA score plot (Figure 3A) with no points overlapped. In order to identify biomarkers for discriminating finger citron samples, variable importance in projection (VIP) scores were calculated for each volatile. It has been reported that variables with value exceeding 1.0 play important roles in the PLS-DA discriminant process [48]. A total of 24 volatiles with VIP > 1.0 were identified (Figure S1C), indicating these were volatiles responsible for clustering and separation of finger citron samples. Among the top 10 compounds, nine were monoterpenoids, including γ-terpinene, α-terpinene, camphene, terpinolene, α-pinene, α-thujene, β-pinene, terpineol, and β-terpineol. Compared to flower and leaf, the concentrations of these monoterpenoids were higher in fruit, and they tended to accumulate during fruit ripening (Figure 2, Table S1). For volatiles mentioned above, terpinene isomers γ-terpinene and α-terpinene were the top two volatiles with high VIP values of 1.78 and 1.76, respectively.
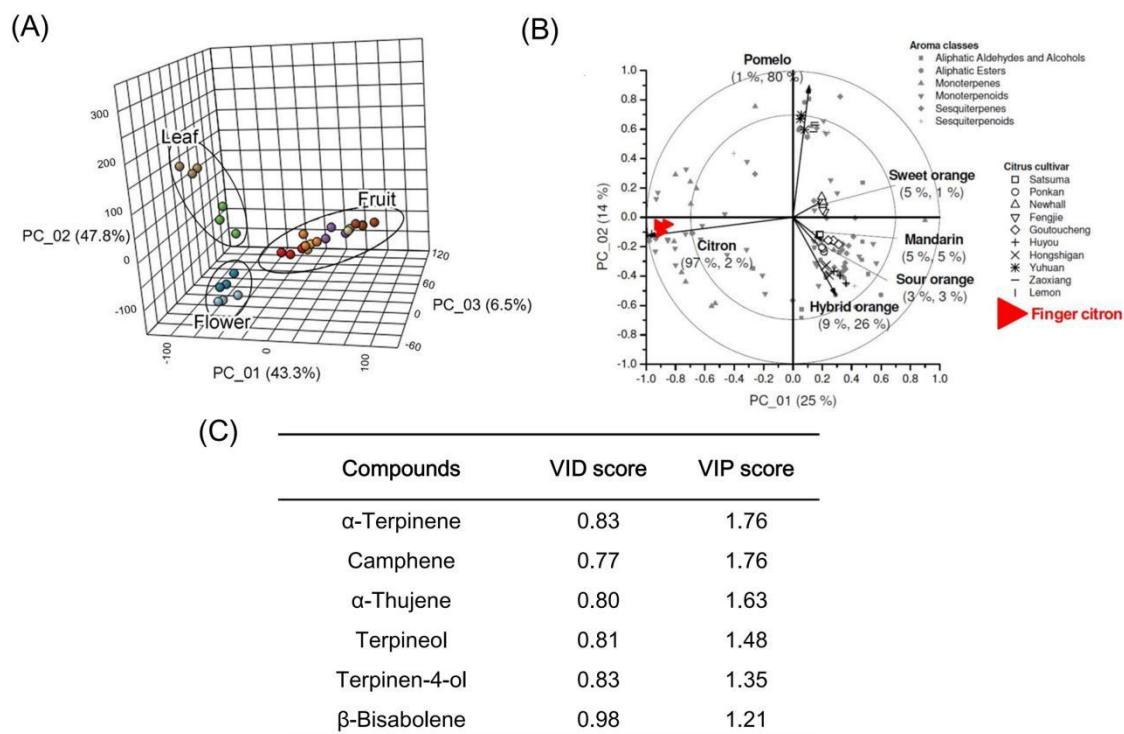

Compounds	VID score	VIP score
α-Terpinene	0.83	1.76
Camphene	0.77	1.76
α-Thujene	0.80	1.63
Terpineol	0.81	1.48
Terpinen-4-ol	0.83	1.35
β-Bisabolene	0.98	1.21

Figure 3. Volatiles ranked by variable identification (VID) and variable importance in projection (VIP) scores. (**A**) partial least squares discriminant analysis (PLS-DA) 3D score plot of finger citron using volatiles as variables. (**B**) PLS bi-plots of eleven *Citrus* varieties from six species. (**C**) Key volatiles ranked by VID and VIP scores. VID scores were calculated using Unscrambler vs 10.1 according to Shen et al.; VIP score was obtained by PLS-DA using online tool MetaboAnalyst 4.0.

Our previous study observed distinct volatile profiles of finger citron fruit compared to various other Citrus species and varieties [25]. PLS analysis was used in order to distinguish the six species by their volatile profiles (Figure 3B). The first two PLS-factors were able to explain 99% variance of citron relative to pummelo, hybrids, sour orange, mandarin, and sweet orange and the clustering results were in agreement with classic Citrus taxonomy. Volatile compounds that have important roles in discriminating citrus fruit samples were identified by Shen et al. and the VID-coefficients were calculated [25]. As an arbitrary threshold, a VID coefficient with absolute value of 0.70 was taken [49]. By comparing the VID-coefficient and VIP values for each volatile compound, 6 volatiles with VID score exceeding 0.70 and VIP score exceeding 1.0 were observed (Figure 3C). These volatiles are suggested to be characteristic compounds of finger citron separating it from other Citrus species (Figure 3).

2.3. A Transcriptome Atlas of Finger Citron

In an attempt to understand the changes in gene expression causing metabolite diversity during development, we sequenced the RNA libraries extracted from flower, leaf and fruit. Information about RNA-Seq quality is summarized in Table S2. The correlation coefficients of tested samples were listed in Figure S2. The correlation of the three biological replicates is above 0.7, with false discovery rate (FDR) value < 0.01 indicating the reliability of sampling. The relatively low correlation between organs indicates a selectively expression pattern of finger citron genes. A total of 1.5G clean data was generated, 95.80% of which was successfully mapped to the reference genome of *C. medica*. We assembled 32,579 genes, which were expressed in finger citron. For functional annotation, unigenes were searched against the Gene Ontology (GO) knowledgebase and the Kyoto encyclopedia of genes and genomes (KEGG). About 76% of the unigenes were assigned to GO terms and 42% of them were annotated with KEGG entries. This information establishes the foundation for further investigation into target genes. Fragments per kilobase of exon model per million mapped reads (FPKM) values were calculated to represent gene expression levels and the gene number and FPKM distribution are summarized in Figure S3. Most genes were expressed in the 0.3~3.57 and 3.57~15 intervals (Figure S3). In general, flowers had the most genes expressed and fruit had the least. As mentioned above, plant terpenoids are synthesized from two independent pathways and the upstream genes and their functions are well studied [2]. Hence, we screened key structural genes involved in terpene biosynthesis by KEGG annotation. The values of their expression fold changes are listed in Table 1.

Table 1. The FPKM fold change (FC) of structure genes in terpenoid synthesis during organ development.

Gene ID	Gene Name	FC (F2/F1)	DEG	FC (L2/L1)	DEG	FC (S6/S1)	DEG
MEP Pathway							
Cm152440	DXS	1.21		0.49	yes, down	0.20	yes, down
Cm165980	DXS	0.97		1.24		0.12	yes, down
Cm165990	DXS	1.42		2.03		1.34	
Cm219000	DXS	1.09		1.72		1.32	
Cm104830	DXS	0.88		2.08	yes, up	1.49	
Cm018790	DXR	0.95		1.25		0.93	
Cm018780	DXR	1.43		0.86		1.22	
Cm191510	MCT	1.39		0.61		0.90	
Cm162030	CMK	1.34		0.88		0.91	
Cm051360	MDS	0.88		1.60		2.53	yes, up
Cm131750	HDS	1.20		2.67	yes, up	1.52	
Cm315300	HDR	0.31		-		-	
Cm315660	HDR	0.24		-		-	
Cm219660	HDR	0.99		0.80		0.42	yes, down
Cm276630	HDR	1.41		2.10	yes, up	0.93	
Cm318240	HDR	1.05		-		0.61	
MVA Pathway							
Cm183050	AACT	0.92		1.03		1.28	
Cm042040	AACT	1.04		0.60		2.90	yes, up
Cm014300	HMGS	1.17		0.82		1.24	
Cm248930	HMGS	0.98		0.38	yes, down	1.47	
Cm197870	HMGR	0.93		0.91		1.42	
Cm212470	HMGR	0.72		0.86		0.66	
Cm122170	HMGR	1.56		0.33	yes, down	5.51	yes, up
Cm268880	MVK	1.23		0.97		0.55	
Cm034040	PMK	1.10		0.52		1.08	
Cm235330	MVD	1.13		0.55		0.65	

Genes whose log2(FC) ≥ 1 and *p* < 0.05 were considered as differentially expressed genes (DEGs). AACT, Acetoacetyl-CoA thiolase; HMGS, HMG-CoA synthase; HMGR, HMG-CoA reductase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; MVD, mevalonate diphosphate decarboxylase; FPPS, farnesyl diphosphate synthase; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; MCT, 4-diphosphocytidyl- 2-C-methyl-D-erythritol synthase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MDS, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase.

Table 1 shows the DEGs in MEP and MVA pathways during organ development. In total, 12 DEGs were identified. No gene transcripts changed significantly during flower development. In the MEP pathway, two DXSs (catalyzing the formation of 1-deoxy-D-xylulose 5-phosphate), Cm152440, and Cm165980 were significantly down-regulated in fruit. DXS/Cm152440 was down-regulated in leaf while another DXS/Cm104830 showed an opposite trend. MDS/ Cm051360 and HDS/ Cm131750 were up-regulated in fruit and flower, respectively. In the MVA pathway, there were four genes differentially expressed: two were down-regulated in leaf and two up-regulated in fruit. During fruit development, the expression level of HMGR, which is considered as the rate limiting enzyme in the MVA pathway [50], was up-regulated by 5.5-fold.

2.4. Changes in Transcript Levels of TPS Gene Family

Terpene synthases catalyze the final pathway steps converting the universal substrates GPP, FPP, and GGPP into terpenoids. There are many members of the TPS family and their complex catalytic mechanisms contribute to the terpenoid diversity in the plant kingdom [4]. A total of 58 *TPS* genes were identified from finger citron based on KEGG annotation. A schematic diagram of TPSs gene structure (Figure S4) shows that finger citron *TPS* gene lengths ranged from about 200 bp to 16,000 bp. Some of these sequences are predicted to be inactive partial fragments in comparison to characterized TPSs in tomato and sweet orange [9,15].

In order to provide an overview of expression profiles of the *TPS* gene family, their transcript levels were presented and visualized as a heat map (Figure 4). As found for volatiles in Figure 2, the *TPS* genes in Figure 4 were also clustered into groups based on their expression patterns, named TPS-Cluster1 to 4. For flower, genes in TPS-Cluster1 were mainly accumulated (Figure 4). The expression of six *TPS* genes (Cm105420, Cm186730, Cm269930, Cm126620, Cm257250, and Cm050840) increased during flower development, while four genes (Cm294730, Cm105550, Cm116860, and Cm257230) decreased. Genes from TPS-Cluster2 and TPS-Cluster3 were predominantly expressed in young and mature leaves, respectively. These two clusters together contain 25 genes, accounting for 38% of *TPS* gene family members. For TPS-Cluster4, genes were mainly expressed in fruit. During fruit development, transcript levels of two *TPS* genes (Cm186720 and Cm269920) increased markedly from S1 to S2 and then remained constant afterwards. High transcript levels of *TPS* genes Cm107350 and Cm285480 were detected in ripe fruit at S6 stage (Figure 4). The spatial and temporal expressing pattern implies the diverse function of TPSs and the regulating mechanism behind terpenoid metabolism, although extremely complex, can now be probed thanks to the developing of modern techniques.

To evaluate expression pattern based on RNA-Seq, RT-qPCR analysis was performed. Four fingers citron TPS genes were randomly selected for gene expression analysis during fruit ripening. As shown in Figure S5, RT-qPCR results matched the expression pattern produced by RNA-Seq. These results indicated that the RNA-Seq data of the present study were accurate and reliable.
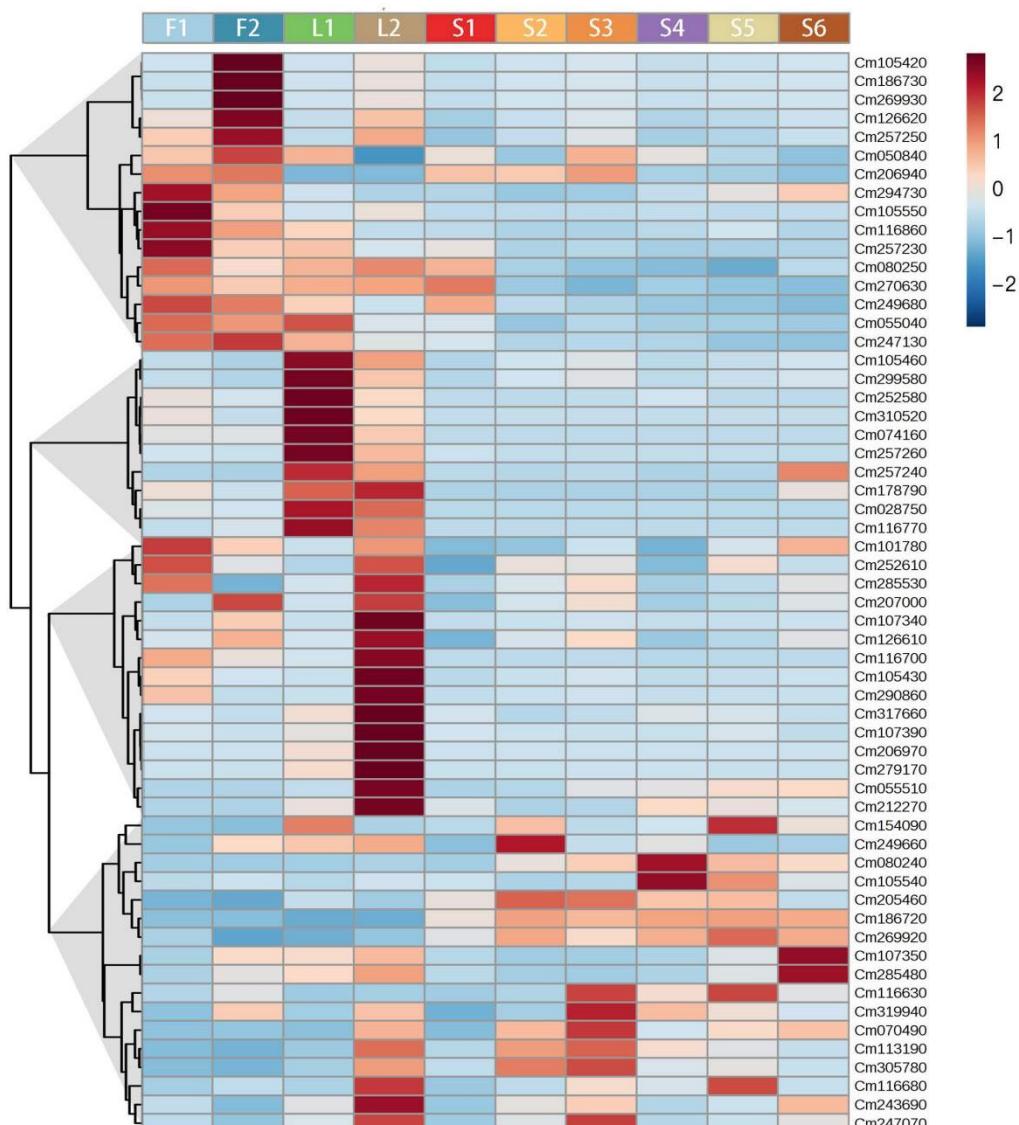
Figure 4. Hierarchical clustering and heat map visualization of changes in *TPS* gene expression level in developing finger citron organs. F1, flower bud; F2, full flower, L1, young leaf; L2, mature leaf; S1–6, fruit developing stages 1–6. The left side of the heat map represents hierarchical clustering based on Pearson correlation. The color scale (−2 to 2) is shown on the right; red represents high content, blue represents low content. Hierarchical clustering and heat map were constructed using MetaboAnalyst 4.0.

2.5. Changes in Transcript Levels of Transcription Factors

The transcript levels of *TPS* genes were regulated by a number of different TFs. So far, six TF families in plants associated with terpenoid metabolism have been identified, including AP2/ERF [25,26,32,33], bHLH [31], MYB [27], NAC [28], WRKY [29], and bZIP [30]. Their gene IDs are summarized in Table S3. Changes in the transcript levels of these six TF families from developing finger citron tissues were analyzed and presented as heat maps (Figure S6). The AP2/ERF family with 251 members was the largest, followed by the NAC family with 184 members and bHLH family with 169 members. As shown in Figure S5, TFs from different families had unique expression patterns.

For the AP2/ERF family, members were grouped into clusters according to expression profiles in developing tissues of finger citron (Figure S6). There were 88 (35%) *AP2/ERFs* mainly expressed in developing fruit, where nine members had the highest expression level in ripe fruit at the S6 stage. For flowers, 36 *AP2/ERFs* had the most abundant transcripts in full flower stage at F2. For finger citron

leaves, higher transcript levels of 18 *AP2/ERFs* were observed in young leave L1 than mature leave L2 (Figure S6).

Among the 169 members in the bHLH family, 58 genes were mainly expressed in fruit, 43 genes in flowers and 68 genes in leaves (Figure S6). For fruit, 11 *bHLHs* had the highest expression level in young fruit at the S1 stage. During flower development, expression of 21 *bHLHs* increased from F1 to F2, while 22 members decreased. Of the *bHLHs* mainly expressed in leaves, 58 genes had higher transcript abundance in young leaves L1 than mature leaves L2 (Figure S6).

A total of 76 bZIP members were identified in finger citron, where 31 (41%) members were mainly expressed in leaves (Figure S6). For MYB family, 58 (37%) of the total 155 family members were enriched in flower tissue. A total of 44 *MYBs* exhibited higher transcript levels in young flower at F1 (Figure S6). We observed that *NAC* genes were enriched in flowers, where 52 (28%) members were expressed mainly in young flowers (Figure S6). Nearly half of the *WRKYs* (27 out of 66 genes) were enriched in fruit tissues. All of these TF expression profiles, together with those of the *TPS* genes, were used for the gene co-expression analysis.

2.6. Analysis of Gene Correlation Network

Both volatiles and genes showed dynamic changes in developing organs of finger citron. Weighted gene co-expression network analysis (WGCNA) was performed to get a better understanding of the relationships between them and their place in the complex signaling networks regulating terpenoid metabolism. A total of 45 gene modules from the dynamic hierarchical tree cut were obtained after a series of matrix transformation procedures and network construction (Figure 5). The largest module was Module Turquoise, containing 1954 genes, and the smallest, 30 genes each, was observed for Module Ivory and Module Floralwhite.
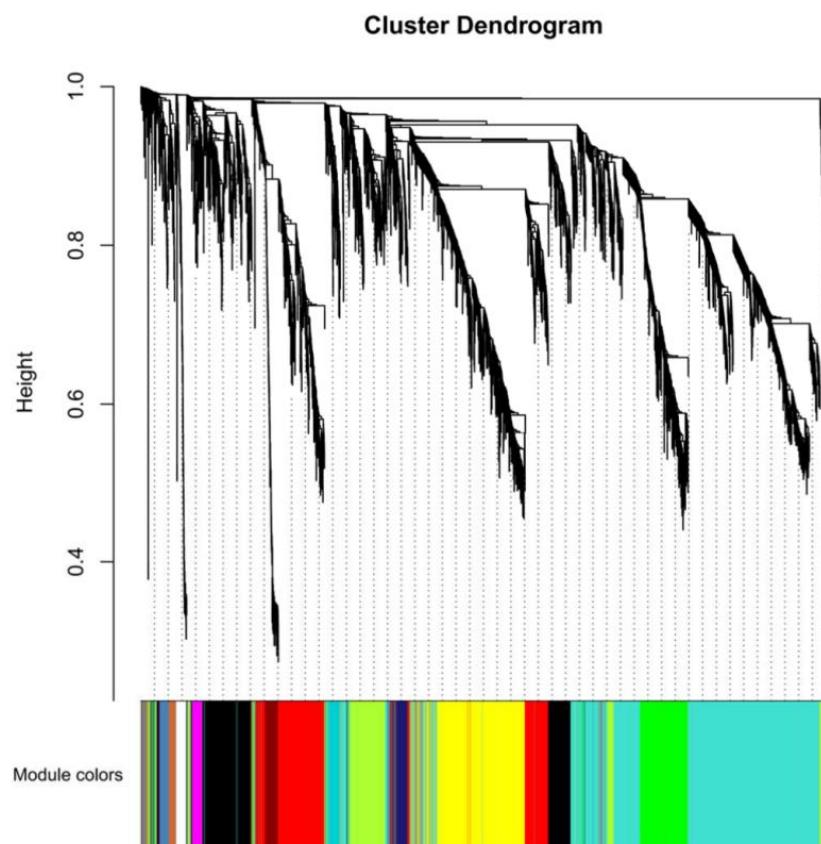

Figure 5. Clustering dendrogram of expressed genes. Gene modules were identified by dynamic hierarchical tree cut and shown in different colors. Height cut = 0.25, minimal module size = 30.

To evaluate the biological importance of the modules, the Pearson correlation between Module eigengenes and external sample traits (in this case, volatile concentrations) [44] were calculated and visualized in Figure 6. The most abundant two components (limonene and γ-terpinene) with the six characteristic components identified in Figure 3C were selected as interesting traits. The terpene synthase gene numbers in each module are shown on the left. In Figure 6, the 58 *TPSs* were assigned to 22 modules according to their transcript profiles. Finger citron *TPSs* occurred in more than twenty modules. This result is predictable because Figure 4 already implied a diversity of *TPS* expression patterns. The modules with the highest number of *TPS* genes were Module Red (11) Module Green (7) and Module Blue (5). In addition to the only sesquiterpene, β-bisabolene, the remaining volatiles showed similar correlation coefficients with a certain module. For example, the correlation coefficients between Module Red and seven monoterpenoids fell in a narrow interval (−0.62 to −0.52) and these monoterpenoids have similar accumulation patterns in finger citron tissues (Figure 2).

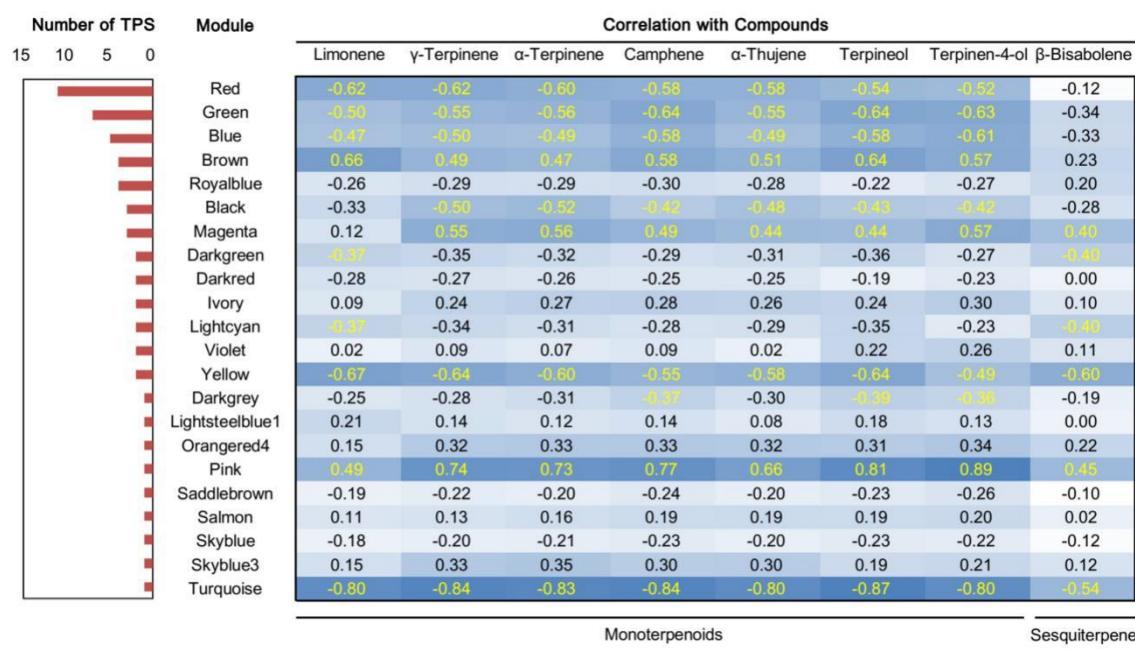|	Limonene	γ-Terpinene	α-Terpinene	Camphene	α-Thujene	Terpineol	Terpinen-4-ol	β-Bisabolene
Red	-0.62	-0.62	-0.60	-0.58	-0.58	-0.54	-0.52	-0.12
Green	-0.50	-0.55	-0.56	-0.64	-0.55	-0.64	-0.63	-0.34
Blue	-0.47	-0.50	-0.49	-0.58	-0.49	-0.58	-0.61	-0.33
Brown	0.66	0.49	0.47	0.58	0.51	0.64	0.57	0.23
Royalblue	-0.26	-0.29	-0.29	-0.30	-0.28	-0.22	-0.27	0.20
Black	-0.33	-0.50	-0.52	-0.42	-0.48	-0.43	-0.42	-0.28
Magenta	0.12	0.55	0.56	0.49	0.44	0.44	0.57	0.40
Darkgreen	-0.37	-0.35	-0.32	-0.29	-0.31	-0.36	-0.27	-0.40
Darkred	-0.28	-0.27	-0.26	-0.25	-0.25	-0.19	-0.23	0.00
Ivory	0.09	0.24	0.27	0.28	0.26	0.24	0.30	0.10
Lightcyan	-0.37	-0.34	-0.31	-0.28	-0.29	-0.35	-0.23	-0.40
Violet	0.02	0.09	0.07	0.09	0.02	0.22	0.26	0.11
Yellow	-0.67	-0.64	-0.60	-0.55	-0.58	-0.64	-0.49	-0.60
Darkgrey	-0.25	-0.28	-0.31	-0.37	-0.30	-0.39	-0.36	-0.19
Lightsteelblue1	0.21	0.14	0.12	0.14	0.08	0.18	0.13	0.00
Orangered4	0.15	0.32	0.33	0.33	0.32	0.31	0.34	0.22
Pink	0.49	0.74	0.73	0.77	0.66	0.81	0.89	0.45
Saddlebrown	-0.19	-0.22	-0.20	-0.24	-0.20	-0.23	-0.26	-0.10
Salmon	0.11	0.13	0.16	0.19	0.19	0.19	0.20	0.02
Skyblue	-0.18	-0.20	-0.21	-0.23	-0.20	-0.23	-0.22	-0.12
Skyblue3	0.15	0.33	0.35	0.30	0.30	0.19	0.21	0.12
Turquoise	-0.80	-0.84	-0.83	-0.84	-0.80	-0.87	-0.80	-0.54

Figure 6. Correlation coefficients between terpene synthase (TPS), module and volatile compound. TPS numbers assigned to each module are list on the left. Correlation coefficients between volatiles and Module eigengenes are presented on the right with a color scale (−1 to 1). Darker blue color represents higher correlation. Coefficients with $p < 0.05$ are colored in yellow.

Module Red, Module Green and Module Blue were large modules, comprising 661, 707, and 1406 nodes, respectively. Connections of genes were hard to distinguish in the co-expression networks constructed on all nodes (Figure S7). So, to get a better view of genes strongly correlated to *TPS* genes, subnetworks were extracted from these three modules using TPSs as seed nodes. Subnetworks constructed from Module Red and Module Blue were visualized as Figures 7 and 8, respectively. Although Module Blue is the largest of the three analyzed modules, no TF was identified in its subnetwork (Figure S8). In these subnetworks, two categories of genes which are the main focus of this work: *TPS* genes (red nodes) and transcription factors (blue nodes). Edges in figures indicate the interactions between the node gene and surrounding TFs in a module.

The subnetwork of Module Red contains four *TPSs*: Cm105420, Cm186730, Cm207000, and Cm269930 (Figure 7). Cm186730 was located on the right of the network, while the other three genes gathered in the center. A total of five TFs were observed in this subnetwork. As a member of NAC family, Cm303080 was connected to two *TPSs*, Cm207000 and Cm105420 (Figure 7). These two *TPS* genes were also connected to MYB family member Cm258540. The *TPS* gene Cm105420, transcript levels were also suggested to be related to transcription factor C2H2 family member Cm279030

and G2-like family member Cm127980. For *TPS* gene Cm269930, transcription factor CPP family Cm308600 was connected (Figure 7, Table S4). No transcription factors were observed to be connected to TPS Cm186730. Based the information in Table S4, Cm070050, Cm269580, and Cm264610 were all predicted to be polygalacturonase genes. This result indicates that, terpenoid metabolism and cell wall metabolism appear to be related in this specific module.

There were six *TPSs* (Cm310520, Cm252580, Cm074160, Cm257260, Cm116770, and Cm028750) in the Module Green subnetwork. Four of them clustered together, including Cm252580, Cm074160, Cm257260, and Cm116770. *TPS* gene Cm257260 was connected to four TFs, including MYB family Cm028420, bHLH family Cm110180 and Cm027020, and ARR-B family Cm022190 (Table S5). For the other two *TPS* genes, Cm028750 was localized at the right edge, while Cm310520 on the left connected to 27 genes. Cm3102520, whose expression was suggested to be related to six TFs (Figure 8). These six TFs included MIKC_MADS family Cm111830, ARR-B family Cm022190, AP2 family Cm215520, ARF family Cm061930, bHLH family Cm227880, and HD-ZIP family Cm175400 (Table S5). Functions of other genes were also investigated (Figure S5). There were genes predicted to participate in secondary metabolism: Cm112960 and Cm107930. Cm093530 and Cm023630 was related to hormone signal transduction.
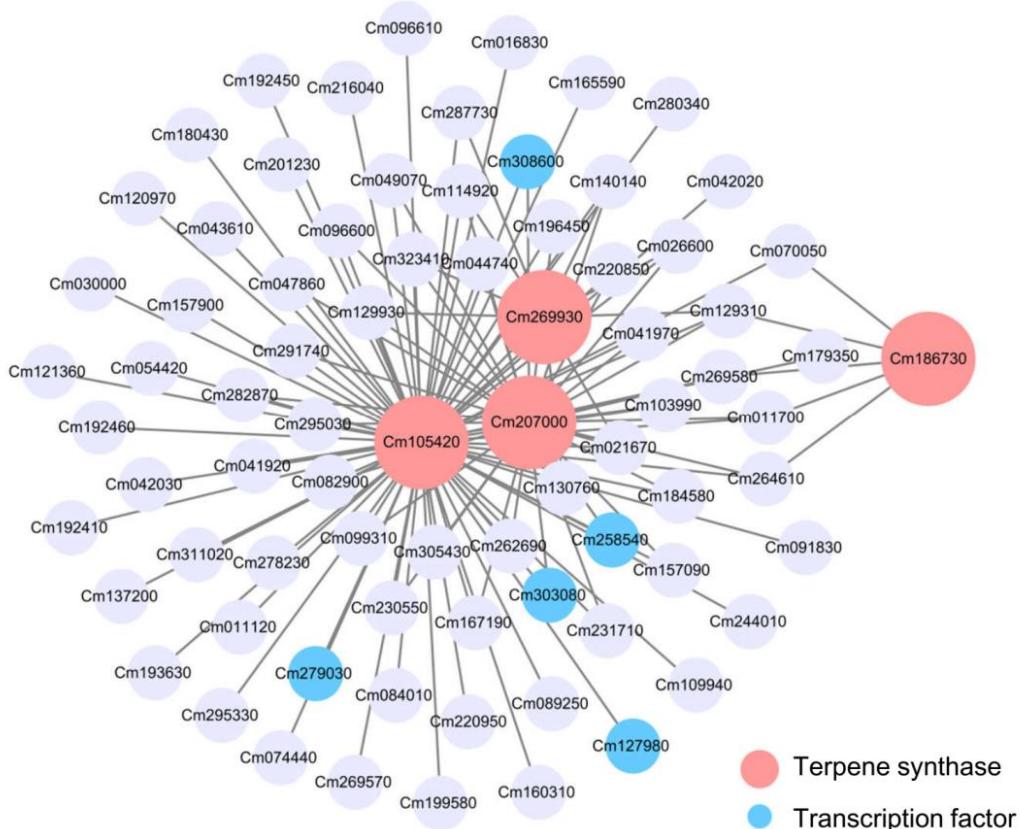
Figure 7. Gene co-expression subnetwork of Module Red. Network was reconstructed by edge weight cutoff = 0.35 and visualize by cytoscape. Gene IDs and annotations in Module Red subnetwork were listed in Table S4.
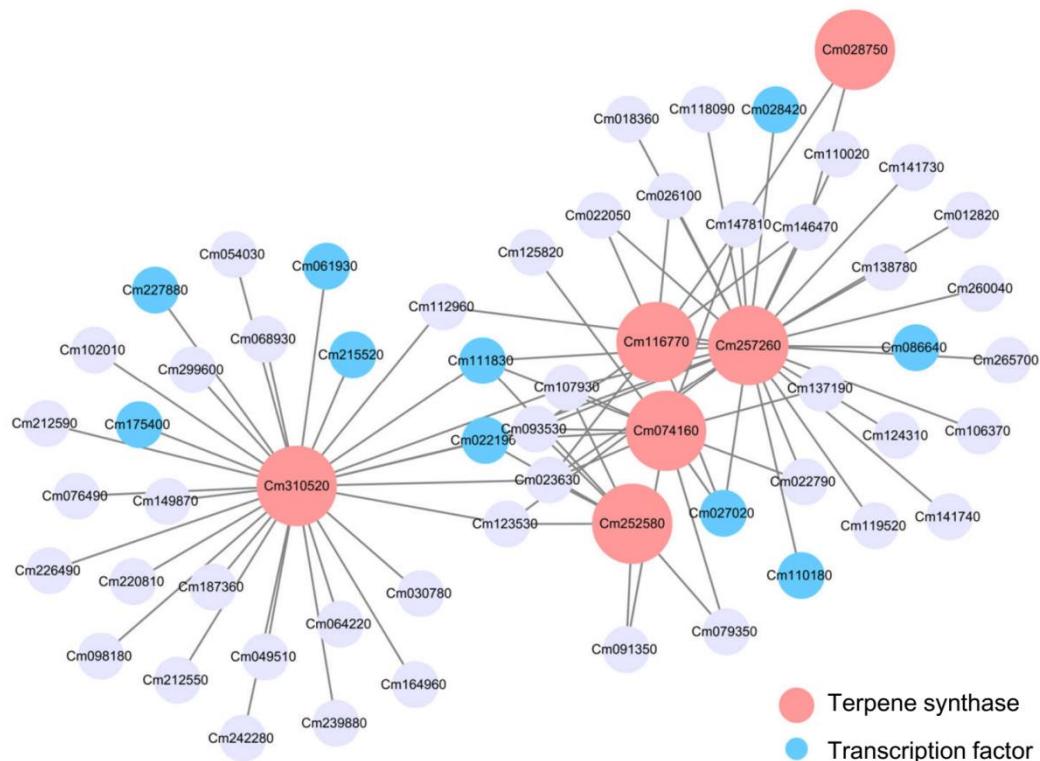
Figure 8. Gene co-expression subnetwork of Module Green. Network was reconstructed by edge weight cutoff = 0.35 and visualize by cytoscape. Gene IDs and annotations in Module Green subnetwork were listed in Table S5.

2.7. Phylogenetic Tree Analysis of Finger Citron TPS

The functions of the TPSs remained unknown, so we constructed a phylogenetic tree based on their amino acid sequences (Figure 9) together with the Citrus TPS reported by Alquézar et al. In Figure 9, *TPS* genes in the phylogenetic tree were highlighted by red dots. Finger citron *TPS* genes were assigned to four subfamilies: 20 in TPS-a, 18 in TPS-b, 2 in TPS-e/f, and 4 in TPS-g. Genes in the TPS-a subfamily are predicted to catalyze the formation of sesquiterpenes [12]. There were 20 finger citron *TPS* belong to this subfamily. Genes in TPS-b usually encode proteins that function as monoterpene synthases. TPS-g is a clade closely related to TPS-b and there is evidence that monoterpenes, sesquiterpenes and diterpenes can be formed by genes from the TPS-g clade [10,51]. Despite their diverse functions, they share the common feature that all their products are acyclic. There are two finger citron TPSs that belong to the e/f clade. To get a more precise prediction of finger citron TPS function, sequence identities compared with characterized *TPS* genes from other *Citrus* species were calculated and are summarized in Figure S9. Since the correlated volatiles to Module Red and Module Green were monoterpenoids (Figure 6), attention was paid to genes in TPS-b and g subfamilies.

The extracted Subnetwork of Module Red contains 4 *TPS* genes. Cm269930 and Cm186730 shared 97.66% identity and they both belonged to TPS-b subfamily (Figure 9). Cm269930 was the hub gene, highly connected node, in Module Red. Sequence analysis showed that Cm269930 shared high amino acid sequence identity with *C. unshiu* γ-terpinene synthase BAD27258 (72.78%) and *C. limon* γ-terpinene synthase AAM53943 (71.01%) (Figure S9). Low sequence identity (29.26%–40.00%) was observed between Cm105420 and *Citrus TPSs* (Figure S9). Module Green contains 6 *TPSs*. TPS family member Cm310520 was ruled out for the phylogenetic analysis because of its inadequate length. Three members (Cm028750, Cm074160, and Cm257260) belong to TPS-b family. Sequence analysis showed that Cm257260 had 90.7% identity with *C. limon* (+)-limonene synthase AAM53944 and *C. jambhiri* limonene synthase BAF73932 (Figure S9). Cm116700 and Cm252580 belong to TPS-a family (Figure 9) and the Cm116700 sequence was 95.02% identical with *C. sinensis* sesquiterpene synthases

Cs4g12350 (Figure S9). Cm028750 and Cm074160 sequence identities were less than 50% homolgous to *Citrus TPSs* sequences. For three *TPS* genes from Module Blue, Cm206970 and Cm279170 belong to TPS-a subfamily, while Cm107390 belong to TPS-g subfamily (Figure 9). Sequence analysis revealed that Cm206970 and Cm279170 had 96.17% and 97.76% identity with *C. unshiu* linalool synthase BAP75561, respectively (Figure S9). For Cm107390, sequence identities to characterized *Citrus TPSs* were less than 45%. These results indicated possible roles of finger citron TPSs from three Modules (Red, Green, and Blue) in biosynthesis of volatile terpenoids, although their functions need to be further investigated.
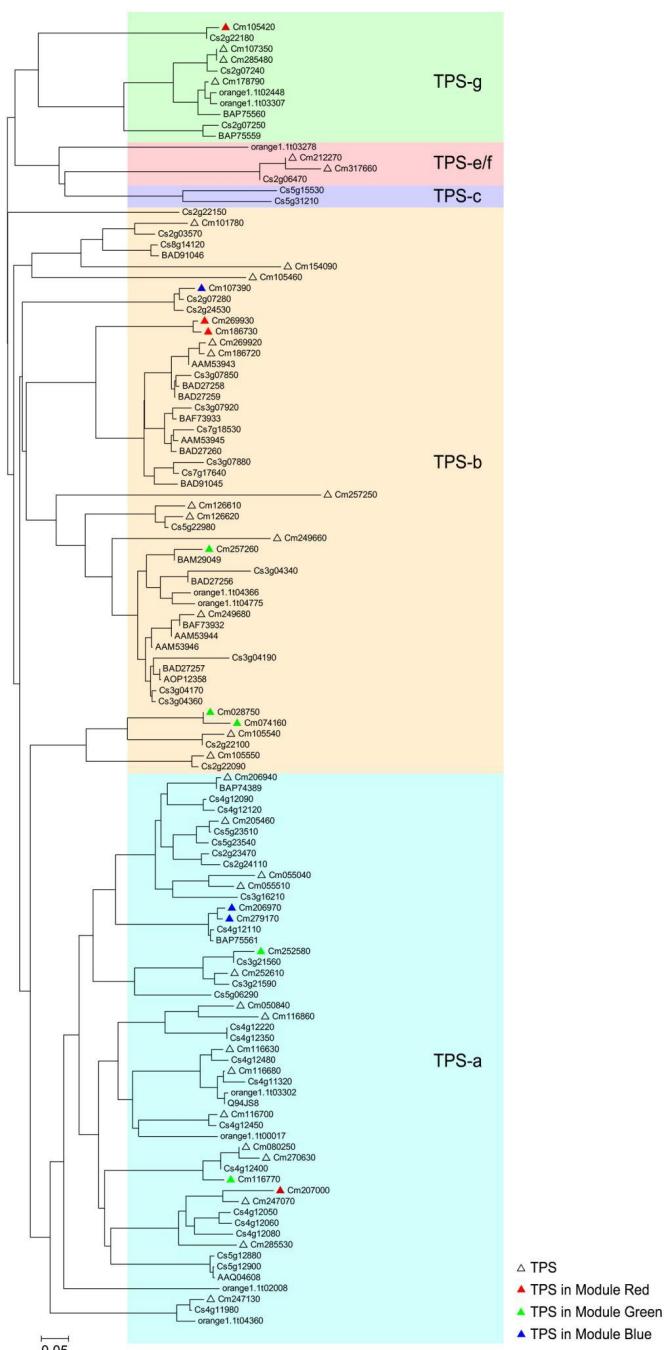
Figure 9. Phylogeny tree analysis of *Citrus* TPSs. Triangles represent finger citron TPSs, and filled colors indicate their location in modules shown in Figure 5. TPSs accession number of sweet orange and other *Citrus* species were reported by Alquézar et al. [15]. Sequence alignment was performed by ClustalX. Phylogeny tree was visualized by FigTree.

3. Discussion

Terpenoids are important components of citrus fruit volatiles. Finger citron has unique aroma, but the molecular mechanism of terpenoid biosynthesis remains unclear. In order to identify the characteristic compounds in this cultivar, we profiled the volatiles in 10 tissues from different finger citron organs. In total, 62 volatile chemicals were characterized, most of which were terpenoids. Consistent with previous studies in finger citron [38,39], the most abundant volatile is limonene followed by γ-terpinene. Variations in volatile compositions were also observed in different organs of *Citrus limon* and other *Citrus* species [52–54]. In the clustering heat map (Figure 2), we observed that most components are differentially expressed in different organs. The main component limonene, decreased when fruit matured and a similar pattern in finger citron was also observed by Wu et al. [40]. A decrease in limonene during fruit ripening was also detected in *Satsuma mandarin* [55,56]. We identified six characteristic components in finger citron (Figure 3). Apart from β-bisabolene, the other five volatiles are cyclic monoterpenoids. They are enriched in fruit and accumulated during fruit ripening. These components are suitable biomarkers that could be used to distinguish finger citron organs and finger citron extracts from those of other *Citrus* species.

We identified 58 members of the finger citron TPS family. This family size is smaller than in grapevine (152) and sweet orange (95), but similar to rice (57) and *Sorghum bicolor* (48) [12]. The largest clade of *TPS* genes found in finger citron, with 20 members, is the TPS-a clade. In sweet orange, TPS-a is also the largest clade [15]. In other plants where the TPS family has been analyzed at genome wide level, TPS-a clade is also the largest clade, including sweet orange [15], tomato [8], apple [11], and grape [10]. It has been suggested that more recent *TPS* gene duplications are indicated by an additional TPS-a clade [11]. Sesquiterpenes are major products produced by TPS-a enzymes, and the high number of TPS-a clade was suggested to be related to their possible role in plant developmental responses, such as defense [15]. RNA-Seq has the ability to simultaneously detect whole gene expression levels, and the availability of such a complete transcriptome profile has been a powerful tool to obtain insights into the molecular mechanisms underlying different conditions. In the present study, transcript levels of finger citron *TPSs* were widely detected in different finger citron organs during development. Similar observations were also made for apple and tomato TPS family in developing organs [9,11], suggesting important biological roles of TPS family members in plants.

To understand the regulation of finger citron terpenoid biosynthesis, expressed genes detected by RNA-Seq were connected by weighted gene co-expression network analysis (WGCNA). This provided a network made of nodes (genes) and edges (connections). Connections in the network were obtained based on gene co-expression data. This strategy has been applied in mining potential targets genes and transcription factors in plants [45–47]. In the present study, the top three modules enriched for *TPS* genes were selected for further analysis. Briefly, thirteen *TPS* genes were screened based on WGCNA, including four from Module Red (Cm269930, Cm207000, Cm105420, and Cm186730), six from Module Green (Cm028750, Cm116770, Cm257260, Cm074160, Cm252580, and Cm310520) and three from Module Blue (Cm206970, Cm107390, and Cm279170). Moreover, a total of 15 TFs were identified, including five from Module Red (Cm127980, Cm258540, Cm279030, Cm303080, and Cm308600) and 10 from Module Green (Cm022190, Cm027020, Cm028420, Cm061930, Cm086640, Cm110180, Cm111830, Cm175400, Cm215520, and Cm227880). These TFs consisted of 12 families, including ARR-B, bHLH, MYB, ARF, MYB, MIKC-MADS, HD-ZIP, AP2/ERF, G2-like, C2H2, NAC, and CPP. Compared to previous studies [25,26,28,57–59], our study revealed several novel TF families, ARR-B, ARF, MIKC-MADS, HD-ZIP, G2-like, and CPP, were associated with *TPSs* expression and volatile terpenoid formation. In the future, more experimental evidence is required to confirm the role of these candidate transcription factors.

Regarding to *TPS* genes identified based on WGCNA, six belong to TPS-a family, five belong to TPS-b and one belongs to TPS-g (Cm107390). TPS-a members were associated with sesquiterpene formation, TPS-b contribute to monoterpene biosynthesis, TPS-g is associated with monoterpene and sesquiterpene formation. Therefore, these identified TPS may have important roles in production

of volatile terpenoids in finger citron. The *TPS* gene Cm310520 sequence is too short to include in the phylogenetic tree construction. Among these thirteen *TPS* genes, four (Cm105420, Cm028750, Cm074160, Cm107390) members had low sequence identity (<50%) to *Citrus TPS* genes with known functions. Relatively, high sequence identities (up to 97.76%) with characterized *Citrus TPS* genes were observed for eight TPSs of finger citron. As the final enzymatic step of MVA and MEP pathways, TPSs are responsible for the direct synthesis of terpenoids. However, in most case expression level of these *TPS* genes do not have a linear correlation with their product contents. The uncertainty arises for two main reasons. Firstly, a considerable number of TPSs are multi-product enzymes which produce several volatiles from a single substrate. Secondly, the replication and evolution of the TPS family has produced functioning isozymes express differently in time and space [4,20,60]. We have previously characterized a sesquiterpene synthase, *CmTPS1*, which has substrate selectivity, catalyzing the formation of 15 sesquiterpenes from FPP, with the main product bicyclogermacrene, followed by aromadenrene and elixene [24]. On the other hand, genes from separate clades have also evolved towards a similar function [8,12]. Therefore, more experiments are necessary to test functions of these 13-candidate finger citron TPS, such as overexpression in *E.coli* for enzymatic analysis in vitro and stable transformation for function analysis in vivo.

4. Materials and Methods

4.1. Plant Materials

Finger citron (*C. media* L. var. *sarcodactylis*) trees were grown in a commercial orchard in Jinhua, Zhejiang, China. Fruit at different developmental stages (S1-S6) were sampled at fixed time intervals: 55 days after bloom (DAB), 85 DAB, 115 DAB, 151 DAB, 182 DAB, and 212 DAB (Figure 1). Changes in fruit weight are shown in Figure 1B. Throughout the experimental period, average fruit weight increased steadily from 68.32 g to 260.77 g. Flower and leaf samples were collected at two developmental stages from the same orchard, with three biological replicates with three fruits each and 15 flowers and 20 leaves each [24]. Fruit of eleven *Citrus* cultivars from six species were collected at the commercial ripened stages as described by Shen et al. [25]. Materials were cut into pieces immediately after collecting, frozen in liquid nitrogen and stored at −80 °C.

4.2. Volatile Collection

Volatile compounds of three finger citron tissues were extracted using head space solid phase microextraction (SPME), using detailed procedures according to Xu et al. [24]. Frozen samples were ground into a fine powder under liquid nitrogen before use. In a 4 mL vial, 0.1 g tissue powder, 1 mL saturated NaCl solution, 10 μL 1-hexanol (0.1%, *v/v*) as internal standard were added. The vial was sealed with a PTFE-silicon septum and vortexed vigorously to mix well. The samples were incubated at 42 °C for 30 min, then extracted with an SPME fiber (50/30 DVB/CAR/PDMS) (Supelco Co., Bellefonte, PA, USA) for 30 min.

4.3. GC-MS Analysis

An Agilent 7890N gas chromatograph (GC) equipped with an HP-5 column (30 m × 0.25 × 0.25 μm, J & W Scientific, Folsom, CA, USA) coupled to an Agilent 5975C Network Mass Selective Detector (MS, inert XL MSD with triple-axis detector) were used for detection of volatiles. After extraction, volatile compounds were detached from the SPME fiber for 5 min at 250 °C by inserting the fiber into the GC injector. The GC oven temperature was started at 40 °C, increased to 70 °C at 3 °C /min, then to 130 °C at 1 °C/min, and finally ramped up to 230 °C at 15 °C/min. Carrier gas was helium with a flow rate at 1.0 mL/min. For detection of bicyclogermacrene, electron impact MS was used in selected ion current mode with an ionization energy of 70 ev. Mass spectra for compound comparisons were acquired over the range 35–350 *m/z* mass units. Volatile compounds were identified and checked by several methods. Chemicals were identified by comparing the electron ionization mass

spectra of individual compounds with the standard spectrum of compounds from the NIST8.0 Mass Spectral Library (http://chemdata.nist.gov/). Kovats retention index of each volatile was calculated and compared to the same volatile identified by other researchers. Finally, when authentic standards were available, the retention time of each volatile was checked under exactly the same condition. Kovats retention index was calculated by injecting a C7–C21 linear alkane mixture standard, based on their molecular structure and retention time. Adding an internal standard as added as a reference helped in semi-quantitative determination of compounds.

4.4. Gene Expression Analysis by RNA Sequencing and Real-Time qPCR

For sequencing, total RNA was isolated from the ten finger citron tissues according to the protocol described by Xu et al. [24]. RNA was digested by Turbo-DNase (Ambion, Foster City, CA, USA). RNA quality was evaluated on a NanoPhotometer® spectrophotometer (IMPLEN, München, Germany). The integrity and quantity of total RNA were assessed using a BioAnalyzer (Agilent Technologies, La Jolla, CA, USA). RNA-Seq was performed by LC-bio (Hangzhou, China) on the Illumina HiSeq 4000 platform. Raw reads obtained from RNA-Seq were pre-processed, adapters were trimmed; low-quality and shorter reads were removed. The *Citrus medica* genome was selected as reference for finger citron. Clean reads were aligned to the *C. medica* genome available online (http://citrus.hzau.edu.cn/orange/) by Hisat. Q20, Q30 and GC contents of the clean data were calculated. The expression profile of assembled gene was attached in Table S6. Functional annotation of genes was obtained from two public databases: KEGG (The Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology). Gene expression level was evaluated by fragments per kilobase of exon model per million mapped reads (FPKM). Three different RNA isolations were used as replicates for libraries construction and RNA sequencing.

To confirm the expression profiles obtained from the transcriptome data, the expression levels of four TPS genes were randomly selected for real-time qPCR (RT-qPCR) analysis. Total RNA was extracted as mentioned above. The first strand cDNA was synthesized using iScript™ cDNA synthesis kit (Bio-Rad, Hercules, CA, USA). The reaction system of qPCR was constructed using reaction mix SsoFast™ EvaGreen® Supermix (Bio-Rad, Hercules, CA, USA) in a volume of 20 μL, according to the manufacturer's instructions. The qPCR reaction program on CFX96TM Real-time system (Bio-Rad, Hercules, CA, USA) was set as follows: 95 °C for 3 min, 45 cycles of 95 °C for 10 s and 60 °C for 30 s, with a final melting curve step from 65 °C to 95 °C. Melting curve analysis for each pair of primers was performed to evaluate the PCR product specificity. Moreover, the PCR products were cloned into pMD18-T vectors (TaKaRa, Dalian, China) and sequence analysis confirmed the correct amplicons produced from each pair of primers. The primers were designed at the 3'UTR and sequences were listed in Table S7. At least three different RNA isolations and cDNA syntheses were used as replicates for the RT-qPCR. The relative expression levels were calculated as $2^{\triangle CT}$ against internal reference citrus actin [24].

4.5. Sequence Analysis and Phylogenetic Tree Construction

Candidate *TPS* genes were screened by KEGG annotations and their sequence was confirmed by BLAST. *TPS* gene structure was visualized by DOMAIN/MOTIF Pattern Drawers from TBtools (v 0.58, South China Agricultural University, Guangzhou, China). A phylogenic tree of CmTPS was generated by ClustalX v 1.83 (Conway Institute UCD Dublin, Ireland) and visualized by MEGA7 v 7.0.14 [61]. *Citrus medica* Transcription factors were obtained by BLAST, taking *Citrus sinensis* transcription factors as reference. Reference sequences were downloaded from the public database PlanTFDB (http://planttfdb.cbi.pku.edu.cn/). TFs were divided into families based on their motifs. Sequence identity (%) of finger citron TPS with other *Citrus* species was calculated on MAFFT (https://www.ebi.ac.uk/Tools/msa/mafft/).

4.6. Gene Co-Expression Network Construction

Weighted gene co-expression network analysis was performed using R package [44]. A pairwise Pearson correlation matrix was created and further transformed into a weighted matrix. A topological overlap matrix was constructed with a threshold power of eight. A dynamic tree cut procedure (merge Cut Height = 0.25, min module size = 30) was performed to identify similar modules in a hierarchical tree [62]. In total, 45 modules were identified. The module sizes ranged from 30 to 1954. In the network, gene connectivity was based on the edge weight (ranging from 0 to 1) determined by the topology overlap measure. Genes with the highest connectivity inside a module were defined as hub genes. The module eigengene E is defined as the first principal component of a given module. Networks were visualized by Cytoscape software v 3.6.1 [63].

4.7. Statistical Analysis

Average and standard derivations of chemicals were calculated using Microsoft Excel software, as were several Figures. Figures were produced and merged by Origin Pro v 8.0 (Origin Lab Corporation, USA). Duncan's test was applied to detect the significant differences between groups at significant level of 0.05 using SPSS v 19.0 (SPSS Inc., Armonk, NY, USA). Heat map and Partial least squares discriminant analysis (PLS-DA) was carried out by online tool MetaboAnalyst v 4.0 (https://www.metaboanalyst.ca/). Venn diagram was constructed by online tool Venny v 2.1.0 (BioinfoGP Service, Madrid, Spain). The value of variable importance in projection (VIP) was also calculated as a variable to evaluate the contribution of a certain compound. Metabolites whose VIP values exceeding 1.0 were selected for further analysis. The variable identification (VID) coefficients of volatile compounds were calculated using Unscrambler v 10.1 (CAMO Technologies Inc., USA) to identify the most important compounds contributing to sample separation. The VID coefficients were calculated as the correlation coefficient between each original X-variable (volatile compounds) and the Y-variable (*Citrus* species) as predicted by the PLS model. As an arbitrary threshold, a VID coefficient with absolute value of 0.70 was taken [49].

5. Conclusions

Our investigation into finger citron metabolite and transcriptome provides new insights into the biosynthesis of terpenoids. In finger citron, expression of upstream genes in the MVA and MEP pathways did not change much during organ development. Terpenoid diversity was caused by differential TPS expression. Based on PLS models, we identified 6 volatiles differentially accumulated in different finger citron organs. Together with the two volatiles with highest contents, these can be considered as the characteristic components of finger citron and can be used to distinguish finger citron from other *Citrus* species. Our results showed there are 58 members of the finger citron *TPS* gene family. The expression patterns of six TF families with 901 members in finger citron organs were also explored. Finally, we obtained 45 gene modules from WGCNA. Based on the high-connectivity subnetwork of Module Red and Module Green, 13 *TPS*s were suggested to play important roles in finger citron characteristic aroma. In addition, 15 TFs were identified whose locations in the subnetwork implied they function in regulating terpenoid metabolism. Future work should focus on the direct and indirect interaction among these target genes and regulators to elucidate the functioning network that controls terpene production.

Supplementary Materials: The following are available online. Table S1: Content (μg/g FW) of 62 identified volatiles in finger citron; Table S2: Summary of the Finger Citron Transcriptome; Table S3: Summary of TF family gene ID. Table S4: Gene ID and annotations in Module Red. Table S5: Gene ID and annotations in Module Green. Table S6: Primer pairs used in qRT-PCR validation. Table S7: The expression profile of assembled genes in RNA-Seq. Figure S1: Results of permutation tests for PLS-DA using volatiles as variables; Figure S2: Pearson correlations between tested samples based on RNA-Seq data; Figure S3: Expression levels of genes based on RNA-Seq data; Figure S4: Schematic diagram of finger citron *TPS* gene structure. Figure S5: Expression analysis of finger citron TPS genes using RNA-Seq and RT-qPCR. Figure S6: Hierarchical clustering and heat map visualization of changes in transcription factor expression level in developing organs of finger citron; Figure S7:

Module Red and Module green coexpression networks; Figure S8: Gene co-expression subnetwork of Module Blue and gene annotations; Figure S9: Amino acid sequence similarity of finger citron TPSs and functional characterized TPSs from other *Citrus* species.

Author Contributions: Y.X., B.Z. and K.C. conceived and designed the experiments; Y.X. and C.Z. performed the experiments in the laboratory; J.S. provided the materials; Y.X., C.X., and B.Z. analyzed the data, and wrote the article; B.Z., D.G., and K.C. revised the manuscript. All authors read and approved the final article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gershenzon, J.; Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **2007**, *3*, 408–414. [CrossRef]

2. Vranová, E.; Coman, D.; Gruissem, W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu. Rev. Plant Biol.* **2013**, *64*, 665–700. [CrossRef] [PubMed]

3. Kuzuyama, T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Biosci. Biotechnol. Biochem.* **2002**, *66*, 1619–1627. [CrossRef]

4. Degenhardt, J.; Köllner, T.G.; Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **2009**, *70*, 1621–1637. [CrossRef] [PubMed]

5. Holopainen, J.K.; Gershenzon, J. Multiple stress factors and the emission of plant VOCs. *Trends Plant Sci.* **2010**, *15*, 176–184. [CrossRef] [PubMed]

6. Breitmaier, E. Terpenes: Importance, general structure, and biosynthesis. In *Terpenes: Flavors, Fragrances, Pharmaca, Pheromones*; Breitmaier, E., Ed.; Wiley-VCH: Weinheim, Germany, 2006; Volume 1, pp. 1–3.

7. González-Mas, M.C.; Rambla, J.L.; López-Gresa, M.P.; Blázquez, M.A.; Granell, A. Volatile compounds in *Citrus* essential oils: A comprehensive review. *Front. Plant Sci.* **2019**, *10*, 12. [CrossRef]

8. Aubourg, S.; Lecharny, A.; Bohlmann, J. Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genom.* **2002**, *267*, 730–745.

9. Falara, V.; Akhtar, T.A.; Nguyen, T.T.H.; Spytopoulou, E.A.; Bleeker, P.M.; Schauvinhold, I.; Matsuba, Y.; Bonini, M.E.; Schimiller, A.L.; Last, R.L.; et al. The tomato terpene synthase gene family. *Plant Physiol.* **2011**, *157*, 770–789. [CrossRef]

10. Martin, D.M.; Aubourg, S.; Schouwey, M.B.; Daviet, L.; Schalk, M.; Toub, O.; Lund, S.T.; Bohlmann, J. 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 Biol.* **2010**, *10*, 226. [CrossRef] [PubMed]

11. Nieuwenhuizen, N.J.; Green, S.A.; Chen, X.J.; Bailleul, E.J.D.; Matich, A.J.; Wang, M.Y.; Atkinson, R.G. Functional genomics reveals that a compact terpene synthase gene family can account for terpene volatile production in apple. *Plant Physiol.* **2013**, *161*, 787–804. [CrossRef]

12. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **2011**, *66*, 212–229. [CrossRef] [PubMed]

13. Pichersky, E.; Lewinsohn, E. Convergent evolution in plant specialized metabolism. *Annu. Rev. Plant Biol.* **2011**, *62*, 549–566. [CrossRef]

14. Dornelas, M.C.; Mazzafera, P. A genomic approach to characterization of the *Citrus* terpene synthase gene family. *Genet. Mol. Biol.* **2007**, *30*, 832–840. [CrossRef]

15. Alquézar, B.; Rodríguez, A.; Peña, M.D.L.; Peña, L. Genomic analysis of terpene synthase family and functional characterization of seven sesquiterpene synthases from *Citrus sinensis*. *Front. Plant Sci.* **2017**, *8*, 1481. [CrossRef] [PubMed]

16. Shimada, T.; Endo, T.; Fujii, H.; Hara, M.; Ueda, T.; Kira, M.; Omura, M. Molecular cloning and functional characterization of four monoterpene synthase genes from *Citrus unshiu* Marc. *Plant Sci.* **2004**, *166*, 49–58. [CrossRef]

17. Shimada, T.; Endo, T.; Fujii, H.; Hara, M.; Omura, M. Isolation and characterization of (E)–β–ocimene and 1,8 cineole synthases in *Citrus unshiu* Marc. *Plant Sci.* **2005**, *168*, 987–995. [CrossRef]

18. Shimada, T.; Endo, T.; Fujii, H.; Omura, M. Isolation and characterization of a new d-limonene synthase gene with a different expression pattern in *Citrus unshiu* Marc. *Sci. Hortic.* **2005**, *105*, 507–512. [CrossRef]

19. Shimada, T.; Endo, T.; Rodríguez, A.; Fujii, H.; Nakano, M.; Sugiyama, A.; Shimizu, T.; Peña, L.; Omura, M. Isolation and characterization of germacrene A synthases gene in *Citrus unshiu* Marc. *Sci. Hortic.* **2012**, *145*, 102–108. [CrossRef]

20. Shimada, T.; Endo, T.; Fujii, H.; Rodríguez, A.; Peña, L.; Omura, M. Characterization of three linalool synthase genes from *Citrus unshiu* Marc. and analysis of linalool-mediated resistance against *Xanthomonas citri* subsp. *citri* and *Penicilium italicum* in citrus leaves and fruits. *Plant Sci.* **2014**, *229*, 154–166. [CrossRef]

21. Lücker, J.; Tamer, M.K.E.; Schwab, W.; Verstappen, F.W.A.; Plas, L.H.W.V.D.; Bouwmeester, H.J.; Verhoeven, H.A. Monoterpene biosynthesis in lemon (*Citrus limon*): cDNA isolation and functional analysis of four monoterpene synthases. *Eur. J. Biochem.* **2010**, *269*, 3160–3171. [CrossRef]

22. Sharon–Asa, L.; Shalit, M.; Frydman, A.; Bar, E.; Holland, D.; Or, E.; Lavi, U.; Lewinsohn, E.; Eyal, Y. *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. *Plant J.* **2003**, *36*, 664–674. [CrossRef] [PubMed]

23. Maruyama, T.; Ito, M.; Honda, F. Molecular cloning, functional expression and characterization of (E)-β-farnesene synthase from *Citrus junos*. *Biol. Pharm. Bull.* **2001**, *24*, 1171–1175. [CrossRef] [PubMed]

24. Xu, Y.Y.; Wu, B.P.; Cao, X.M.; Zhang, B.; Chen, K.S. *Citrus* CmTPS1 is associated with formation of sesquiterpene bicyclogermacrene. *Sci. Hortic.* **2017**, *226*, 133–140. [CrossRef]

25. Shen, S.L.; Yin, X.R.; Zhang, B.; Xie, X.L.; Jiang, Q.; Grierson, D.; Chen, K.S. CitAP2.10 activation of the terpene synthase *CsTPS1* is associated with the synthesis of (+)-valencene in 'Newhall' orange. *J. Exp. Bot.* **2016**, *67*, 4105–4115. [CrossRef] [PubMed]

26. Li, X.; Xu, Y.Y.; Shen, S.L.; Yin, X.R.; Klee, H.; Zhang, B.; Chen, K.S. Transcription factor CitERF71 activates the terpene synthase gene CitTPS16 involved in the synthesis of *E*-geraniol in sweet orange fruit. *J. Exp. Bot.* **2017**, *68*, 4929–4938. [CrossRef] [PubMed]

27. Reddy, V.A.; Wang, Q.; Dhar, N.; Kumar, N.; Venkatesh, N.; Rajan, C.; Panicker, D.; Sridhar, V.; Mao, H.Z.; Sarojam, R. Spearmint R2R3-MYB transcription factor MsMYB negatively regulates monoterpene production and suppresses the expression of geranyl diphosphate synthase large subunit (MsGPPS.LSU). *Plant Biotechnol. J.* **2017**, *15*, 1105–1119. [CrossRef]

28. Nieuwenhuizen, N.J.; Chen, X.Y.; Wang, M.Y.; Matich, A.J.; Perez, R.L.; Allan, A.C.; Green, S.A.; Atkinson, R.G. Natural variation in monoterpene synthesis in kiwifruit: Transcriptional regulation of terpene synthases by NAC and ETHYLENE-INSENSITIVE3-like transcription factors. *Plant Physiol.* **2015**, *167*, 1243–1258. [CrossRef] [PubMed]

29. Xu, Y.H.; Wang, J.W.; Wang, S.; Wang, J.Y.; Chen, X.Y. Characterization of GaWRKY1, a cotton transcription factor that regulates the sesquiterpene synthase gene (+)-δ-cadinene synthase-A1. *Plant Physiol.* **2004**, *135*, 507–515. [CrossRef]

30. Miyamoto, K.; Nishizawa, Y.; Minami, E.; Nojiri, H.; Yamane, H.; Okada, K. Overexpression of the bZIP transcription factor OsbZIP79 suppresses the production of diterpenoid phytoalexin in rice cells. *J. Plant Physiol.* **2015**, *173*, 19–27. [CrossRef]

31. Hong, G.J.; Xue, X.Y.; Mao, Y.B.; Wang, L.J.; Chen, X.Y. Arabidopsis MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *Plant Cell* **2012**, *24*, 2635–2648. [CrossRef]

32. Yu, Z.X.; Li, J.X.; Yang, C.Q.; Hu, W.L.; Wang, L.J.; Chen, X.Y. The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol. Plant* **2012**, *5*, 353–365. [CrossRef] [PubMed]

33. Skinem?Li, L.; Hao, X.L.; Liu, H.; Wang, W.; Fu, X.Q.; Ma, Y.N.; Shen, Q.; Chen, M.H.; Tang, K.X. Jasmonic acid-responsive AabHLH1 positively regulates artemisinin biosynthesis in *Artemisia annua*. *Biotechnol. Appl. Biochem.* **2019**. [CrossRef]

34. Wang, X.; Xu, Y.T.; Zhang, S.Q.; Cao, Y.; Huang, Y.; Cheng, J.F.; Wu, G.Z.; Tian, S.L.; Chen, C.L.; Liu, Y.; et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature Genet.* **2017**, *49*, 765–774. [CrossRef] [PubMed]

35. Liao, F.L.; Wang, Y.J.; Chen, M.G.; Chen, W.R.; Zong, Y.; Li, Y.Q.; Yang, L.; Han, X.X.; Guo, W.D. Low transcription of *CmsIAA9* in the basal pistil is related to parthenocarpic fruiting of fingered citron (Foshou). *Mol. Breed.* **2017**, *37*, 101. [CrossRef]

36. Brigand, J.P.; Nahon, P. Gastronomy and the citron tree (*Citrus medica* L.). *Int. J. Gastro. Food Sci.* **2016**, *3*, 12–16. [CrossRef]

37. Zhang, W.Q. Chapter 37 *Citrus medica* L. var. *sarcodactylis* Swingle (foshou, finger citron). In *Dietary Chinese Herbs: Chemistry, Pharmacology and Clinical Evidence*; Liu, Y.Z., Wang, Z.M., Zhang, J.Z., Eds.; Springer: Friesach, Austria, 2015; pp. 327–331.

38. Lou, Z.X.; Chen, J.; Yu, F.H.; Wang, H.X.; Kou, X.R.; Ma, C.Y.; Zhu, S. The antioxidant, antibacterial, antibiofilm activity of essential oil from *Citrus medica* L. var. *sarcodactylis* and its nanoemulsion. *LWT-Food Sci. Technol.* **2017**, *80*, 371–377. [CrossRef]

39. Peng, C.H.; Ker, Y.B.; Weng, C.F.; Peng, C.C.; Huang, C.N.; Lin, L.Y.; Peng, R.Y. Insulin secretagogue bioactivity of finger citron fruit (*Citrus medica* L. var. *Sarcodactylis* Hort, Rutaceae). *J. Agric. Food Chem.* **2009**, *57*, 8812–8819. [CrossRef]

40. Wu, Z.; Li, H.; Yang, Y.; Zhan, Y.; Tu, D.W. Variation in the components and antioxidant activity of *Citrus medica* L. var. *sarcodactylis* essential oils at different stages of maturity. *Ind. Crop. Prod.* **2013**, *46*, 311–315. [CrossRef]

41. He, Y.F.; Jin, X.L. Review of research progress in *Citrus medica* with reference to prediction of development prospects. *Econ. For. Res.* **2001**, *4*, 41–43. (In Chinese)

42. Jing, L.; Lei, Z.T.; Zhang, G.W.; Pilon, A.C.; Huhman, D.V.; Xie, R.J.; Xi, W.P.; Zhou, Z.Q. Metabolite profiles of essential oils in citrus peels and their taxonomic implications. *Metabolomics* **2015**, *11*, 952–963. [CrossRef]

43. Zhang, H.P.; Xie, Y.X.; Liu, C.H.; Chen, S.L.; Hu, S.S.; Xie, Z.Z.; Deng, X.X.; Xu, J. Comprehensive comparative analysis of volatile compounds in citrus fruits of different species. *Food Chem.* **2017**, *230*, 316–326. [CrossRef] [PubMed]

44. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [CrossRef] [PubMed]

45. Hopper, D.W.; Ghan, R.; Schlauch, K.A.; Cramer, G.R. Transcriptomic network analyses of leaf dehydration responses identify highly connected ABA and ethylene signaling hubs in three grapevine species differing in drought tolerance. *BMC Plant Biol.* **2016**, *16*, 118. [CrossRef] [PubMed]

46. Ferreira, S.S.; Hotta, C.T.; Poelking, V.G.D.C.; Leite, D.C.C.; Buckeridge, M.S.; Loureiro, M.E.; Barbosa, M.H.P.; Carneiro, M.S.; Souza, G.M. Co-expression network analysis reveals transcription factors associated to cell wall biosynthesis in sugarcane. *Plant Mol. Biol.* **2016**, *91*, 15–35. [CrossRef] [PubMed]

47. Tai, Y.L.; Liu, C.; Yu, S.W.; Yang, H.; Sun, J.M.; Guo, C.X.; Huang, B.; Liu, Z.Y.; Yuan, Y.; Xia, E.H.; et al. Gene co-expression network analysis reveals coordinated regulation of three characteristic secondary biosynthetic pathways in tea plant (*Camellia sinensis*). *BMC Genom.* **2018**, *19*, 616. [CrossRef] [PubMed]

48. Steingass, C.B.; Jutzi, M.; Müller, J.; Carle, R.; Schmarr, H.G. Ripening-dependent metabolic changes in the volatiles of pineapple (*Ananas comosus* (L.) Merr.) fruit: II. Multivariate statistical profiling of pineapple aroma compounds based on comprehensive two-dimensional gas chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2015**, *407*, 2609–2624. [CrossRef] [PubMed]

49. Vandendriessche, T.; Keulemans, J.; Geeraerd, A.; Nicolai, B.M.; Hertog, M.L.A.T.M. Evaluation of fast volatile analysis for detection of *Botrytis cinerea* infections in strawberry. *Food Microbiol.* **2012**, *32*, 406–414. [CrossRef]

50. Song, A.A.L.; Abdullah, J.O.; Abdullah, M.P.; Shafee, N.; Othman, R.; Tan, E.F.; Noor, N.M.; Raha, A.R. Overexpressing 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) in the lactococcal mevalonate pathway for heterologous plant sesquiterpene production. *PloS ONE* **2012**, *7*, e52444. [CrossRef]

51. Chen, F.; Tholl, D.; D'Auria, J.C.; Farooq, A.; Pichersky, E.; Gershenzo, J. Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *Plant Cell* **2003**, *15*, 481–494. [CrossRef]

52. Flamini, G.; Tebano, M.; Cioni, P.L. Volatiles emission patterns of different plant organs and pollen of *Citrus limon*. *Anal. Chim. Acta* **2007**, *589*, 120–124. [CrossRef]

53. Azam, M.; Jiang, Q.; Zhang, B.; Xu, C.J.; Chen, K.S. *Citrus* leaf volatiles as affected by developmental stage and genetic type. *Int. J. Mol. Sci.* **2013**, *14*, 17744–17766. [CrossRef] [PubMed]

54. Azam, M.; Zhang, B.; Xu, Y.Y.; Xu, C.J.; Chen, K.S. Comparative analysis of flower volatiles from nine *Citrus* at three blooming stages. *Int. J. Mol. Sci.* **2013**, *14*, 22346–22367. [CrossRef] [PubMed]

55. Gao, J.; Wu, B.P.; Gao, L.X.; Liu, H.R.; Zhang, B.; Sun, C.D.; Chen, K.S. Glycosidically bound volatiles as affected by ripening stages of *Satsuma mandarin* fruit. *Food Chem.* **2018**, *240*, 1097–1105. [CrossRef] [PubMed]

56. Wei, G.; Tian, P.; Zhang, F.X.; Qin, H.; Miao, H.; Chen, Q.W.; Hu, Z.Y.; Cao, L.; Wang, M.J.; Gu, X.F.; et al. Integrative analyses of non-targeted volatile profiling and transcriptome data provide molecular insight into VOC diversity in cucumber plants (*Cucumis sativus*). *Plant Physiol.* **2016**, *172*, 603–618. [CrossRef] [PubMed]

57. Goossens, J.; Mertens, J.; Goossens, A. Role and functioning of bHLH transcription factors in jasmonate signalling. *J. Exp. Bot.* **2017**, *68*, 1333–1347. [CrossRef] [PubMed]

58. Matías-Hernández, L.; Jiang, W.M.; Yang, K.; Tang, K.X.; Brodelius, P.E.; Pelaz, S. AaMYB1, and its orthologue AtMYB61, affect terpene metabolism and trichome development in *Artemisia annua* and *Arabidopsis thaliana*. *Plant J.* **2017**, *90*, 520–534. [CrossRef] [PubMed]

59. Rushton, P.J.; Somssich, I.E.; Ringler, P.; Shen, Q.X. WRKY transcription factors. *Trends Plant Sci.* **2010**, *15*, 247–258. [CrossRef]

60. Tholl, D. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **2006**, *9*, 297–304. [CrossRef]

61. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *22*, 1870–1874. [CrossRef]

62. Yip, A.M.; Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinform.* **2007**, *8*, 22. [CrossRef]

63. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

Sample Availability: Samples of the compounds are available from the authors.