Integrative Analyses of Nontargeted Volatile Profiling and Transcriptome Data Provide Molecular Insight into VOC Diversity in Cucumber Plants (*Cucumis sativus*)[1][OPEN]

Guo Wei, Peng Tian, Fengxia Zhang, Hao Qin, Han Miao, Qingwen Chen, Zhongyi Hu, Li Cao, Meijiao Wang, Xingfang Gu, Sanwen Huang, Mingsheng Chen, and Guodong Wang*

State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China (G.We., P.T., F.Z., H.Q., Q.C., Z.H., L.C., M.W., M.C., G.Wa.); University of the Chinese Academy of Sciences, Beijing 100039, China (G.We., P.T., Q.C., Z.H.); and Key Laboratory of Horticultural Crops Genetic Improvement of the Ministry of Agriculture, Sino-Dutch Joint Laboratory of Horticultural Genomics Technology, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China (H.M., X.G., S.H.)

ORCID IDs: 0000-0002-9830-0102 (L.C.); 0000-0001-7773-5150 (X.G.); 0000-0002-8547-5309 (S.H.); 0000-0001-9917-0656 (G.Wa.).

Plant volatile organic compounds, which are generated in a tissue-specific manner, play important ecological roles in the interactions between plants and their environments, including the well-known functions of attracting pollinators and protecting plants from herbivores/fungi attacks. However, to date, there have not been reports of holistic volatile profiling of the various tissues of a single plant species, even for the model plant species. In this study, we qualitatively and quantitatively analyzed 85 volatile chemicals, including 36 volatile terpenes, in 23 different tissues of cucumber (*Cucumis sativus*) plants using solid-phase microextraction combined with gas chromatography-mass spectrometry. Most volatile chemicals were found to occur in a highly tissue-specific manner. The consensus transcriptomes for each of the 23 cucumber tissues were generated with RNA sequencing data and used in volatile organic compound-gene correlation analysis to screen for candidate genes likely to be involved in cucumber volatile biosynthetic pathways. In vitro biochemical characterization of the candidate enzymes demonstrated that TERPENE SYNTHASE11 (TPS11)/TPS14, TPS01, and TPS15 were responsible for volatile terpenoid production in the roots, flowers, and fruit tissues of cucumber plants, respectively. A functional heteromeric geranyl(geranyl) pyrophosphate synthase, composed of an inactive small subunit (type I) and an active large subunit, was demonstrated to play a key role in monoterpene production in cucumber. In addition to establishing a standard workflow for the elucidation of plant volatile biosynthetic pathways, the knowledge generated from this study lays a solid foundation for future investigations of both the physiological functions of cucumber volatiles and aspects of cucumber flavor improvement.

In nature, plants emit a huge diversity of volatile organic compounds (VOCs). These VOCs, acting together with the nonvolatile specialized chemicals, in both direct and indirect interactions, protect plants from various forms of biotic attacks and facilitate plant adaptation to their environments. VOCs are a vital element of a plant's chemotype, playing a central role in plant ecosystems, as they function as ecological signals. To date, more than 1,700 compounds have been categorized as VOCs of plant origin (Knudsen et al., 2006; Dudareva et al., 2013). Based on their biosynthetic origins, the plant VOCs can be divided roughly into three main classes: terpenoids, which are generated from methylerythritol phosphate (MEP) or mevalonate (MVA); phenylpropanoids/benzenoids, which are generated from aromatic amino acids (such as Phe); and alcohols/aldehydes, which are generated from unsaturated fatty acids (such as C18:2 and C18:3 fatty acids) and amino acids (Fig. 1; Dudareva et al., 2013). Although the enzymes that are responsible for the upstream metabolism of plant VOC production have been characterized in considerable detail, the downstream enzymes of VOC biosynthetic pathways remain largely uncharacterized. This is likely due to the structural diversity and the species-specific property of plant VOCs. In plant biology, the functional characterization of the genes encoding the enzymes of VOC production is a

[1] This work was supported by the National Program on Key Basic Research Projects (973 Program grant no. 2012CB113900) and the State Key Laboratory of Plant Genomics of China (grant no. SKLPG2011B0104 to G.Wa.).

* Address correspondence to gdwang@genetics.ac.cn.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Guodong Wang (gdwang@genetics.ac.cn).

G.We., X.G., S.H., M.C., and G.Wa. designed the research; G.We., F.Z., Q.C., Z.H., and L.C. performed the chemical and biochemical analyses; G.We., P.T., H.Q., and M.W. carried out the bioinformatics analysis; H.M. took care of the cucumber plants and performed sample preparation; G.We. and G.Wa. analyzed data and wrote the article, with comments from all of the authors.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.01051

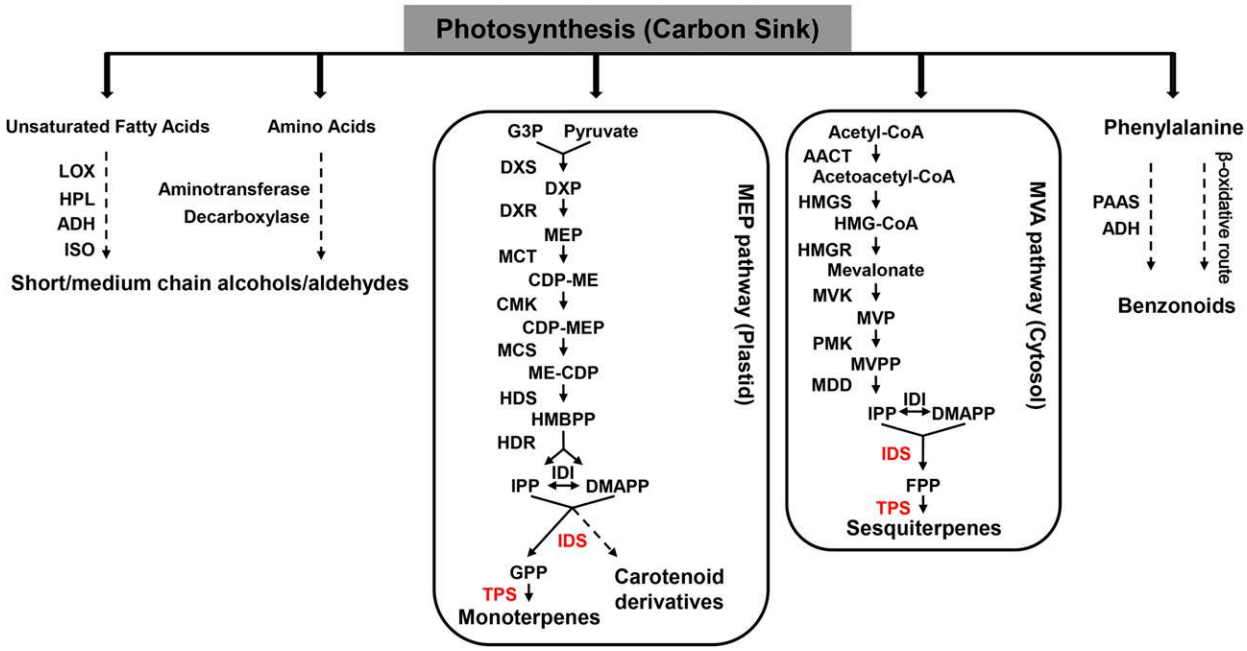

Figure 1. Simplified VOC metabolic network in plant cells. All carbons of plant VOCs are generally derived from plant photosynthesis. The enzymes investigated in this study are highlighted in red, and dashed lines indicate the involvement of multiple enzymatic reactions. For clarity, only biosynthetic pathways for terpenoids, benzenoids, and short/medium-chain alcohols/aldehydes are shown. Abbreviations for enzymes are as follows: AACT, acetyl-CoA acetyltransferase; ADH, alcohol dehydrogenase; CMK, 4-(cytidine 5′-diphospho)-2-*C*-methyl-D-erythritol kinase; DXS, DXP synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; HDS, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase; HMGR, HMG-CoA reductase; HMGS, HMG-CoA synthase; HPL, hydroperoxide lyase; IDI, isopentenyl diphosphate isomerase; IDS, isoprenyl diphosphate synthase; ISO, isomerase; LOX, lipoxygenase; MCS, ME-CDP synthase; MCT, 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase; MDD, mevalonate diphosphate decarboxylase; MVK, mevalonate kinase; PAAS, phenylacetaldehyde synthase; PMK, phosphomevalonate kinase; TPS, terpene synthase. Abbreviations for chemicals are as follows: CDP-ME, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol; CDP-MEP, CDP-ME 2-phosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; G3P, glyceraldehyde 3-phosphate; HMBPP, (*E*)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate; HMG-CoA, hydroxymethylglutaryl-CoA; ME-CDP, 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate; MVP, mevalonate 5-phosphate; MVPP, mevalonate 5-pyrophosphate.

prerequisite for the study of the physiological function(s) of a given plant VOC (Muhlemann et al., 2014). Among these plant VOCs, the biosynthetic pathway of volatile terpenoids, composed mainly by monoterpenes and sesquiterpenes, was investigated extensively during the past three decades. Obviously, the size of the TPS family and the promiscuity of each TPS member are main contributors to the diversity of volatile terpenoids in the plant kingdom (Degenhardt et al., 2009; Chen et al., 2011, and refs. therein). Usually, the TPS enzyme catalyzes GPP and *E,E*-FPP, both of which are trans-prenyl pyrophosphates, to produce monoterpenes and sesquiterpenes, respectively. As shown in Figure 1, IDS, the direct upstream enzyme of TPS, provides the trans-prenyl pyrophosphate for TPSs. Recently, new types of plant IDS were functionally characterized (Sallaud et al., 2009; Schilmiller et al., 2009; Wang et al., 2016).

Previously, studies of plant VOCs have focused mainly on plant tissues that are known to produce or to emit high levels of VOCs, such as plant flowers, fruits, and glandular trichomes (Klee, 2010; Muhlemann et al., 2014; Wang, 2014, and refs. therein). With the rapid development of next-generation sequencing technology, mega-sequencing data mining has become a central component for the discovery of novel genes involved in biosynthetic pathways of interest (Metzker, 2010). This is especially the case for research in plant species lacking a reference genome. Although the amount of such data has expanded to a vast scale, two classical criteria are still used for novel gene elucidation using the reverse genetics strategy: (1) positive correlation between gene expression pattern and the metabolite accumulation pattern; and (2) known biochemical knowledge about the likely functions of enzymes that may catalyze a given enzymatic step of interest. Using a genome-wide correlation analysis, one can expect to identify all of the gene candidates for each enzymatic step of a specific pathway, even the whole metabolic network (Wang, 2014). However, functional characterization of predicted genes is still a laborious procedure in the laboratory.

Fruits and vegetables are important sources of vitamins, minerals, dietary fiber, and antioxidants. Cucumber (*Cucumis sativus*; $2n = 14$; the estimated genome size is 367 Mb) is an economically important vegetable crop; it is well known for its fresh green note flavor when

consumed directly. Mercke et al. (2004) investigated the genes involved in spider mite (*Tetranychus urticae*)-induced volatile production in cucumber leaf discs by using a coresponse strategy. They found that the volatile compound (*Z*)-3-hexenyl acetate and the sesquiterpene (*E*,*E*)-α-farnesene were strongly induced by both spider mite infection and jasmonic acid treatment. Several cDNA contigs were identified as candidate sequences for volatile production based on high Pearson correlation coefficient (PCC) values between volatile emission data and gene expression data. Cucumber (*E*,*E*)-α-farnesene synthase and (*E*)-β-caryophyllene synthase were functionally characterized in vitro (Mercke et al., 2004).A recent study demonstrated that the volatile emission levels from infested cucumber leaves, including (*E*)-β-ocimene, (*E*,*E*)-4,8,12-trimethyltrideca-1,3,7,11-tetraene, and two other unknown compounds, correlated positively with the attraction of *Phytoseiulus persimilis*, a predatory mite species that targets herbivorous spider mites (Kappers et al., 2011). This result suggested that the VOCs can protect cucumber plants indirectly. However, knowledge about the production of volatile from cucumber remains fragmentary.

The available cucumber genome facilitates the functional genomics study of the metabolism in cucumber (Huang et al., 2009; Shang et al., 2014). In this study, we applied combined omics technologies, including solid-phase microextraction (SPME) combined with gas chromatography-mass spectrometry (GC-MS)-based VOC profiling and RNA sequencing (RNAseq) transcriptome data, to globally study volatile production in different tissues of cucumber plants. Our study started with a screen for likely candidate genes (encoding particular classes of enzymes) involved in cucumber volatile biosynthetic pathways based on VOC-gene correlation analysis. We further functionally characterized the genes responsible for terpenoid biosynthesis in cucumber roots, flowers, and fruits. In addition to the mechanistic enzymatic information about volatile biosynthesis, this study also lays a solid foundation for both future investigations about the physiological and ecological functions of cucumber VOCs and future applied efforts to improve cucumber flavor.

RESULTS

The VOC Content of 23 Unique Cucumber Plant Tissues

To deeply understand volatile biosynthesis in cucumber plants, we profiled the volatiles from a total of 23 distinct tissues from cucumber plants (for detailed information, see Fig. 2). To save sampling time and ensure reproducible analytical results, we chose the salt solution (20% NaCl) to prepare analytical samples of different cucumber organs, rather than intact cucumber organs, in this study. Fused silica fibers coated with divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) were selected for volatile collection due to their known properties of good reproducibility and its relatively wide coverage of chemical diversity (Risticevic and Pawliszyn, 2013). A total of 85 volatile compounds were detected in at least one of the 23 cucumber plant tissues (22 volatile compounds were structurally confirmed via comparison with the analytical results for authentic reference standards, and 63 volatile compounds were tentatively identified by spectral matching of analytic spectra with reference mass spectra in the National Institute of Standards and Technology [NIST] 08 database). The list of detected compounds included 27 alcohols/aldehydes, 21 monoterpenoids, 15 sesquiterpenoids, eight benzenoids, four carotenoid derivatives, and 10 unclassified compounds (Supplemental Data Sets S1 and S2). Principal component analysis (PCA) of the profiling data for these 85 VOCs from the 23 cucumber tissues showed that the different tissues from mature plants (12 weeks old) had distinct VOC profiles (roots, young leaves, old leaves, flowers, fruit peel, and fruit flesh). Meanwhile, the same tissue with different developmental stages grouped together (peel of cucumber fruits at various developmental stages [S16, S18, S20, and S22] and the same case in flesh tissues [S17, S19, S21, and S23]), indicating that the same tissue at different developmental stages had similar VOC fingerprints. In contrast, four tissues sampled from 4-week-old seedlings had similar VOC profiles that were distinct from the profiles of mature plants (Fig. 3A). The PCA analysis also clearly indicated that the SPME-GC-MS method used in this study has a satisfactory reproducibility for cucumber VOC profiling; the various biological replicates (n = 2–4) of the same tissue were closely grouped (Fig. 3A).

Clustering analysis showed that most VOCs occurred in a tissue-specific pattern (e.g. aromatic benzenoids were detected in abundance in roots and flowers, while C9 and C6 aldehydes were detected at high levels in fruits and leaves; Fig. 3B). Only eight VOCs, (*E*)-2-hexenal, benzaldehyde, (*Z*)-6-nonenal, azulene, dodecane, 2-methyl-naphthalene, 1-methyl-naphthalene, and geranylacetone, were detected ubiquitously in all 23 tissues sampled (Supplemental Data Set S1), reinforcing the supposition of the tissue specificity for most of the VOCs in cucumber plants. These highly tissue-specific patterns of the VOCs facilitated the screening for candidate genes involved in the biosynthetic pathways of VOCs based on metabolite-gene correlation analysis (Saito et al., 2008; Matsuda et al., 2010). It is noteworthy that most monoterpenes produced in mature roots were in an oxidized form. The possible precursors for these monoterpene oxides and sesquiterpene were deduced and are summarized in Supplemental Figures S1 and S2.

A Cucumber Transcriptome Atlas

In an attempt to deeply understand the biosynthetic pathways of VOCs in cucumber, we further acquired the transcriptome data of 23 cucumber tissues using the RNAseq method. We generated a total of 53G of high-quality base pairs (at least 2G foreach tissue). Bioinformatics analysis showed that approximately

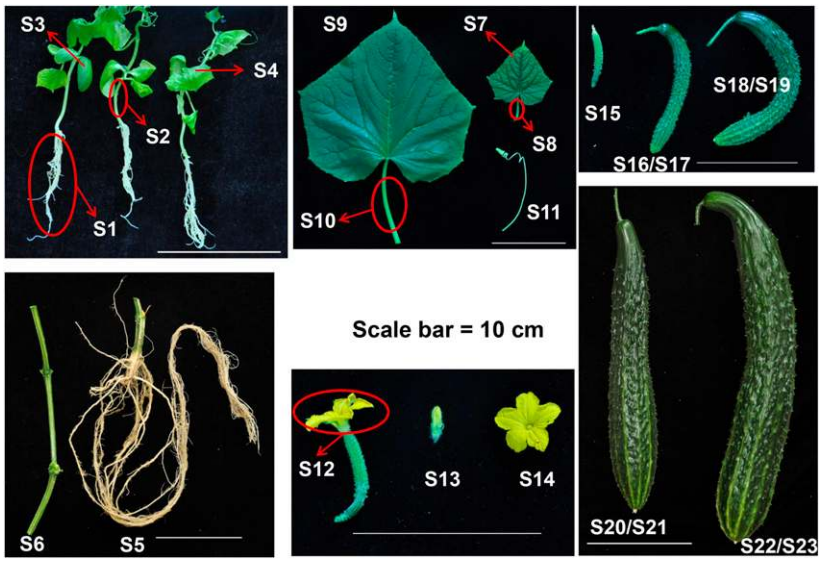

Figure 2. Tissues collected from the cucumber plants for VOC profiling and transcriptome analysis. S1, Roots of 4-week-old seedlings; S2, hypocotyl of 4-week-old seedlings; S3, cotyledon of 4-week-old seedlings; S4, true leaf of 4-week-old seedlings; S5, root; S6, stem; S7, young leaf (the second node from the top of cucumber plants); S8, petiole of young leaf; S9, old leaf (the fifth node from the top of cucumber plants); S10, petiole of old leaf; S11, tendril; S12, female flower; S13, male flower bud; S14, male flower; S15, unfertilized ovary; S16, peel of unfertilized ovary; S17, flesh of unfertilized ovary; S18, peel of 1-week-old fruit; S19, flesh of 1-week-old fruit; S20, peel of 2-week-old fruit; S21, flesh of 2-week-old fruit; S22, peel of 3-week-old fruit; S23, flesh of 3-week-old fruit.

269 million reads were obtained in total, with an average of 12.6 million reads per tissue. Approximately 253 million high-quality reads (94.04% of the total reads) were aligned to the cucumber genome (Supplemental Fig. S3A). Among the aligned reads, 87.87% were uniquely aligned; the remaining 12.13% were mapped to multiple loci. Further analysis showed that the majority of the mapped reads were in exonic regions, compared with approximately one-fifth of the aligned reads in intronic regions and only a small minority of aligned reads in nongene regions (Supplemental Fig. S3B). In total, the expression patterns of 21,788 expressed cucumber genes, whose expression level is higher than 0.5 fragments per kilobase of transcript per million fragments mapped (FPKM) in at least one tested tissue, are summarized in Supplemental Data Set S3. The average transcriptional level of all 21,788 unigenes (there are 26,682 predicted genes in the cucumber genome; Huang et al., 2009) in each cucumber tissue ranged from 67 to 152 FPKM, and 46.59% of the expressed genes ranged from 10 to 100 FPKM (Fig. 4A). In contrast to the highly tissue-specific patterns detected for the cucumber VOCs, we found that only a small fraction (6.26%) of the expressed genes (a total of 1,365 genes; defined by Shannon entropy [H] ≤ 2; for details, see "Materials and Methods") showed tissue-specific expression patterns (Fig. 4B). Consistent with the tissue specificity of plant specialized metabolites, these tissue-specific genes were suggested to be involved in specialized metabolism (including the VOC biosynthesis) in cucumber.

In Silico-Based Screening of the Candidate Genes for VOC Production in Cucumber

To generate a list of candidate genes likely to be involved in cucumber VOC biosynthesis, we tried to integrate the transcriptome data and the VOC profiling data from the 23 tissues by correlation analysis. However, the mismatched magnitude of the transcriptome data (21,788 genes) and the VOC data (85 VOCs) can be assumed to result in a high number of false-positive hits. To reduce the ratio of false positives, we selected 9,914 expressed cucumber unigenes that were assigned to the catalytic category (including all enzyme genes involved in metabolism) within the molecular function category of a Gene Ontology (GO:0003824) analysis (Supplemental Fig. S4). More than 50,000 edges, including VOC-VOC correlations (PCC $r >$ 0.8) and VOC-gene correlations ($P < 0.01$), were identified in the correlation analysis (Supplemental Data Set S4; Supplemental Fig. S5). It is clear that there were almost certainly many false-positive correlations in this correlation analysis, as only a limited number of enzymatic steps (genes) would actually occur in the biosynthesis of main VOCs (Fig. 1). It is noteworthy that the correlation coefficients in this study may be underestimated, given that the emission level of a particular plant VOC is affected by many factors beyond the biosynthesis, such as various biochemical modifications of VOC in plant cells (Schwab et al., 2015; Widhalm et al., 2015). Despite these limitations, the correlation analysis in this study provided a data-mining resource that was useful for the rapid screening of genes likely to be involved in cucumber VOC biosynthetic pathways. The candidate genes, determined by the combination of known knowledge and the PCC value of the VOC-gene correlation, for cucumber VOC biosynthesis are summarized in Table I. To test the robustness of the VOC-gene correlation in this study, we further functionally characterized the selected TPSs and IDSs that we had surmised were likely responsible for the terpene diversity (Table I).

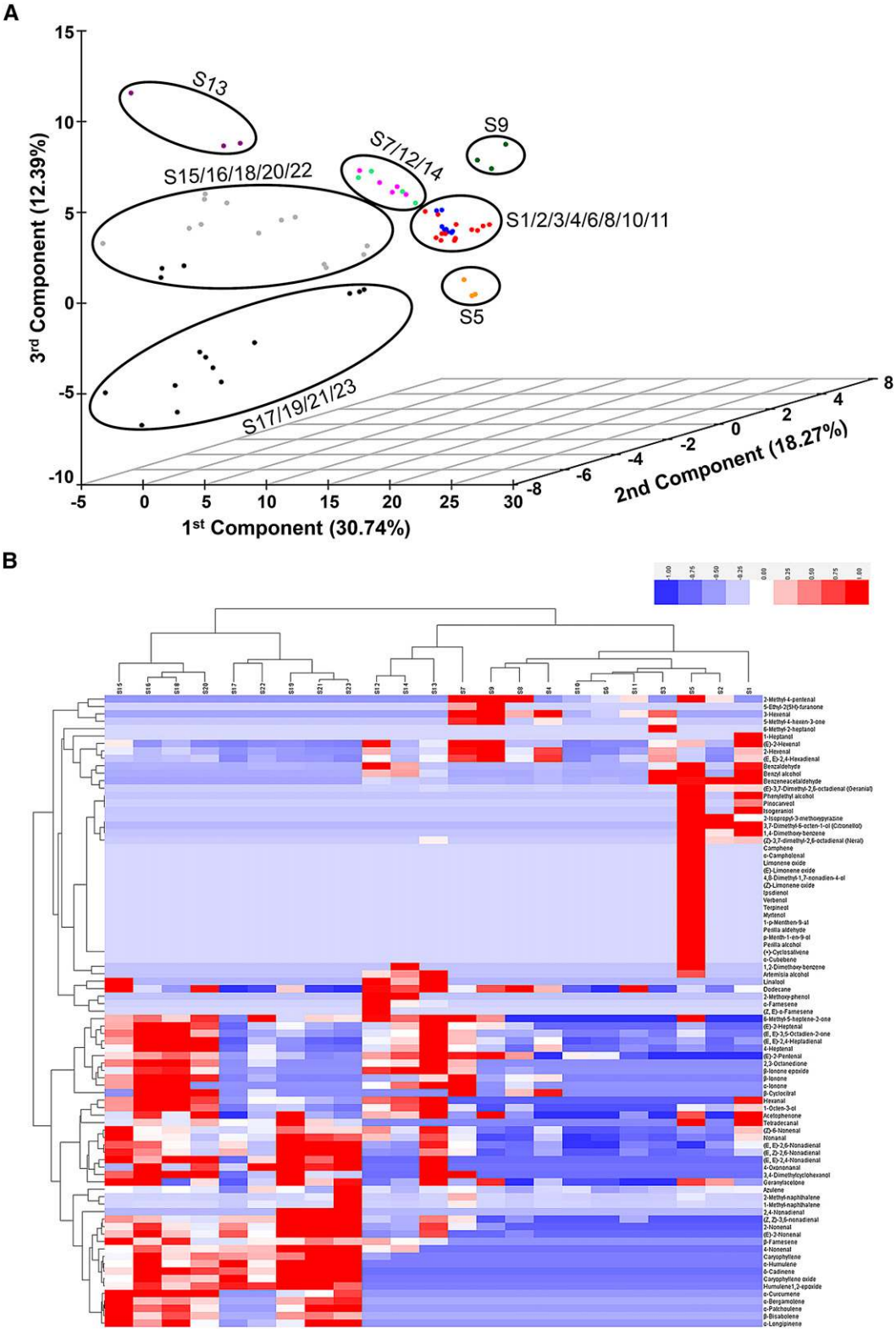

Figure 3. PCA and clustering analysis of cucumber VOCs. A, PCA of VOC profiles among 23 tissues of cucumber plants ($n = 70$). PC1 (30.7% variance), PC2 (18.3% variance), and PC3 (12.4% variance) were used to generate the scores plot. For detailed identification of samples S1 to S23, see the legend to Figure 2. Samples of seedlings (S1–S4), root (S5), stem/petiole (S6, S8, S10, and S11), leaf (S7 and S9), flower (S12–S14), fruit peel (S16, S18, S20, and S22), and fruit flesh (S15, S17, S19, S21, and S23) are marked, respectively, in red, brown, blue, green, purple, gray, and black. B, Heat map of VOC-VOC correlations generated from

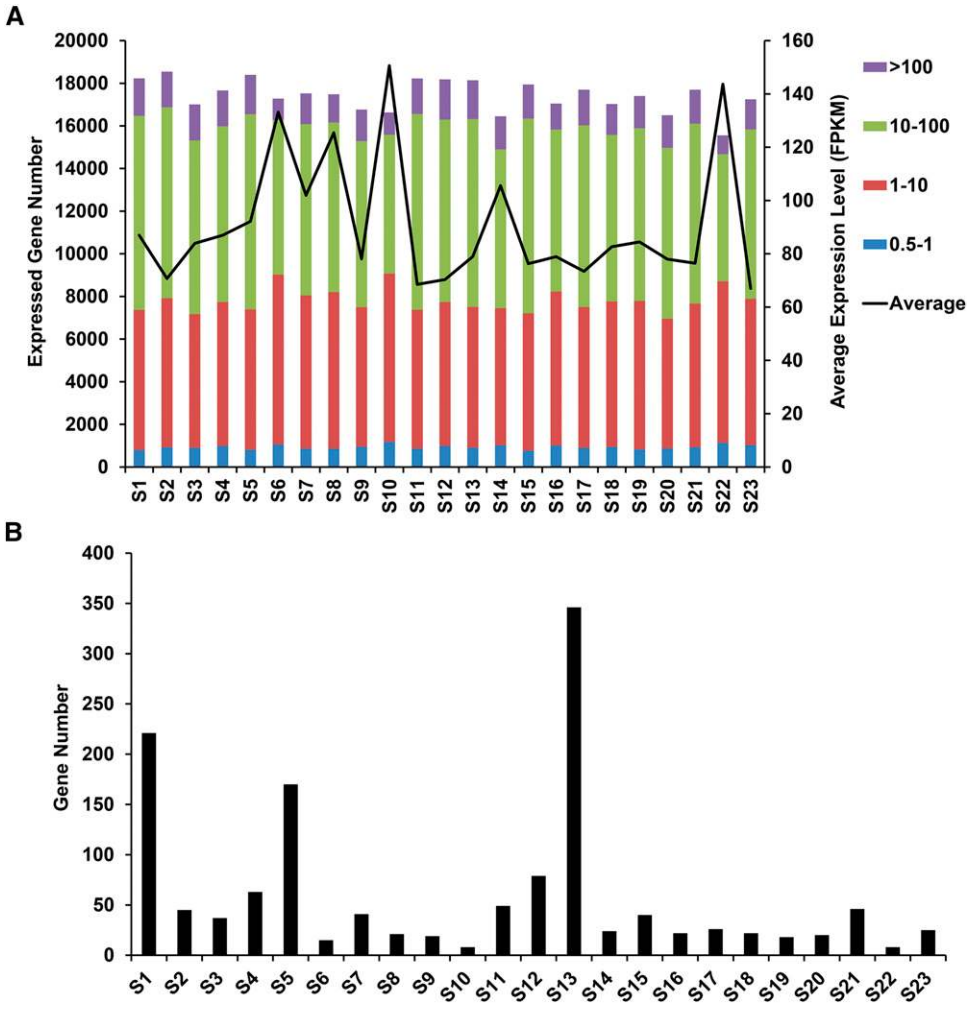

Figure 4. Global and tissue-specific expression patterns of cucumber genes in the 23 sampled tissues. A, Global expression levels of cucumber genes based on RNAseq data from the 23 sampled cucumber tissues. B, Tissue-specific expression of genes distributed among the 23 sampled cucumber tissues ($H \leq 2$).

Characterization of Terpenoid Metabolism-Related Enzymes in Cucumber

According to the VOC profiling results of the sampled cucumber tissues, we initially tried to explain the observed terpene diversity at the molecular genetic level (a total of 36 volatile terpenes were detected; Fig. 2). Unlike the tissue specificity of VOCs in cucumber plants, no genes of the MEP pathway (a total of 14 genes, covering eight enzymatic steps, from 1-deoxy-D-xylulose 5-phosphate synthase to isopentenyl diphosphate isomerase) or the MVA pathway (a total of 13 genes, covering seven enzymatic steps, from acetoacetyl-CoA thiolase to isopentenyl diphosphate isomerase, and two FPP synthases) showed obvious tissue specificity (Supplemental Data Sets S5 and S6). It is well documented that IDSs, together with the directly downstream enzyme, TPSs, are the two primary factors for the skeletal diversity of plant terpenoid (Chen et al., 2011; Jia and Chen, 2016; Wang et al., 2016). We found 24 *TPS* genes (the encoded peptides greater than 200 amino acids) by searching the cucumber genome; thus, they were designated as *CsaTPS01* through *CsaTPS24* in this study (Supplemental Data Set S7). It is noteworthy that the TPS-g (*TPS01–TPS04*), TPS-b (*TPS05–TPS11*), and TPS-a (*TPS13–TPS22*) subfamilies were localized in tandem on chromosomes 1, 2, and 3, respectively (Supplemental Fig. S6). It is known that *TPS* gene clusters formed via tandem duplication are common in flowering plants (Falara et al., 2011; Hofberger et al., 2015). Compared with the midsize *TPS* gene family, there is a relatively small family for the IDS genes in the cucumber genome, including two polyprenyl pyrophosphate synthase genes, two FPP synthase genes, and four geranyl(geranyl) pyrophosphate

Figure 3. (*Continued.*)
the profiling data of the 23 cucumber tissues. VOCs were grouped in tissues, and each rectangle represents the value (in bins; see key) of a correlation between tissues.

Table I. *Representative candidate genes likely involved in cucumber VOC production, determined by VOC-gene Pearson correlation analysis*

VOC Classification	VOC Name	Candidate Gene Identifier	Functional Annotation	PCC
Terpenoid	Total monoterpenoid	*Csa7G211090*	Small subunit of heteromeric G(G)PPS	0.59[a]
	Terpineol	*Csa3G039850*	Terpene synthase	0.99[a]
		Csa2G299330	Terpene synthase	0.99[a]
		Csa2G299920	Terpene synthase	0.99[a]
		Csa2G427840	Terpene synthase	0.99[a]
		Csa3G042380	Terpene synthase	0.99[a]
		Csa1G039830	Cytochrome P450	0.99[a]
		Csa3G015360	Cytochrome P450	0.99[a]
		Csa1G618890	Cytochrome P450	0.98[a]
	Cyclosativene	*Csa3G039850*	Terpene synthase	0.99[a]
		Csa2G299330	Terpene synthase	0.99[a]
		Csa2G299920	Terpene synthase	0.99[a]
		Csa2G427840	Terpene synthase	0.99[a]
		Csa3G042380	Terpene synthase	0.99[a]
	Aromadendrene	*Csa3G040850*	Terpene synthase	0.78[a]
	Linalool	*Csa1G068570*	Terpene synthase	0.37
		Csa1G066550	Terpene synthase	0.25
Alcohol/aldehyde	(*E*,*Z*)-2,6-Nonadienal	*Csa4G288070*	Lipoxygenase	0.75[a]
		Csa6G424000	Lipoxygenase	0.58[a]
		Csa7G075590	Cytochrome P450	0.50[b]
	(*E*)-2-Hexenal	*Csa2G023940*	Lipoxygenase	0.72[a]
		Csa2G024440	Lipoxygenase	0.69[a]
		Csa7G041960	Cytochrome P450	0.70[a]
Benzenoid	Benzeneacetaldehyde	*Csa3G611340*	Tyr decarboxylase	0.27
	Benzaldehyde	*Csa5G318390*	CoA ligase	0.89[a]
		Csa6G319770	Enoyl-CoA hydratase	0.35
		Csa3G133140	3-Ketoacyl-CoA thiolase	0.33
Carotenoid derivative	β-Ionone epoxide	*Csa1G435760*	Carotenoid cleavage dioxygenase	0.57[a]

[a]$P < 0.01$ (Student's *t* test, two tails). [b]$P < 0.05$ (Student's *t* test, two tails).

synthase [G(G)PPS]-like genes (Table I; Supplemental Data Set S8).

The expression patterns of all of these candidate genes (five *TPS* genes and one GGPPS-like gene; Table I) were verified with quantitative RT-PCR (qRT-PCR; Fig. 5). The subcellular localization experiments showed that TPS01, TPS11, and four GGPPS-like proteins were clearly plastid localized (Fig. 6), which mostly agreed with the subcellular localization and signal peptide predictions obtained using ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/) and WoLF PSORT (http://www.wolfpsort.org; Supplemental Data Sets S7 and S8).

Characterization of Root-Specific TPSs

The expression of five *TPS* genes with root-specific expression (CsaTPS07 encoded by *Csa2G299330*, CsaTPS11 encoded by *Csa2G299920*, CsaTPS12 encoded by *Csa2G427840*, CsaTPS14 encoded by *Csa3G039850*, and CsaTPS18 encoded by *Csa3G042380*) that were predicted to be involved in volatile terpene biosynthesis in cucumber root tissue was analyzed via qRT-PCR. The open reading frames of *CsaTPS11*, *CsaTPS12*, and *CsaTPS14* encode peptides of 600, 565, and 562 amino acids, respectively. CsaTPS11 belongs to the TPS-b subfamily and has a predicted plastid-targeting peptide of 47 amino acids at its N terminus (0.506, predicted with TargetP 1.1 software), which was experimentally confirmed (Fig. 6). CsaTPS12 and CsaTPS14 belong to the TPS-a subfamily; these proteins share 56.7% amino acid identity. We did not obtain the open reading frames of *CsaTPS07* or *CsaTPS18*, likely owing to their relatively low transcription level in cucumber root tissue (Supplemental Data Set S7).

A truncated form of *CsaTPS11* and the full open reading frame of *CsaTPS12*/14 were subcloned into the pMAL-c2x vector and expressed in *Escherichia coli*. The purified recombinant proteins were used for TPS assays. GC-MS analysis of enzymatic products showed that all three of these root-specific TPSs catalyzed GPP to monoterpenes and FPP to sesquiterpenes. Among them, CsaTPS11 produced α-pinene, sabinene, and β-pinene as primary products, with other minor monoterpenes [β-thujene, camphene, β-pinene, α-phellandrene, α-terpinen, limonene, eucalyptol, γ-terpinen, and *p*-mentha-1,4(8)-diene, listed here according to the elution order from the gas chromatography analysis] also produced when using GPP as a substrate. It should be noted that limonene (peak 9) and eucalyptol (peak 10) were coeluted in our GC-MS method (Fig. 7A; Supplemental Fig. S7). CsaTPS11 produced α-himachalene, β-sesquiphellandrene, and three other uncharacterized sesquiterpenes when using FPP as a substrate.

Both CsaTPS12 and CsaTPS14 produced similar monoterpene profiles [β-pinene, limonene, and

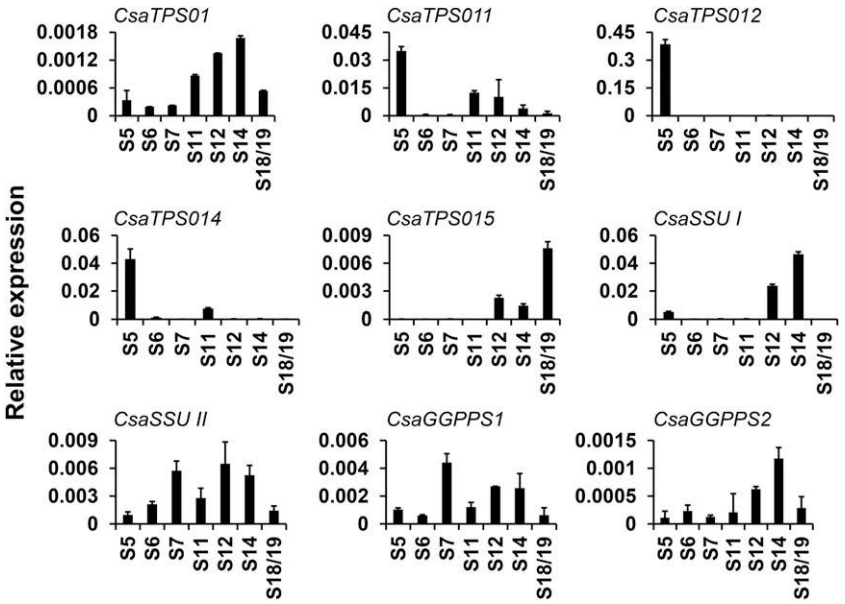

Figure 5. qRT-PCR analysis of the transcript levels of five TPS genes and four IDS genes in seven representative cucumber tissues. Transcript levels are expressed relative to the expression of the gene encoding the ubiquitin extension protein (GenBank accession no. AY372537). Error bars represent the SD of three independent experiments. S5, Root; S6, stem; S7, young leaf; S11, tendril; S12, female flower; S14, male flower; S18/19, 1-week-old fruit (flesh + peel).

p-mentha-1,4(8)-diene], with the exception that CsaTPS12 produced three additional monoterpenes when using GPP as a substrate (α-pinene, camphene, and β-pinene). However, CsaTPS12 and CsaTPS14 produced different sesquiterpene profiles when using FPP as a substrate: trans-α-bergamotene, cis-α-bergamotene, trans-β-farnesene, β-bisabolene, aromadendrene, α-patchoulene, and α-bisabolene were detected in the CsaTPS12 reaction system, while CsaTPS14 produced only cyclosativene and α-cubebene (Fig. 7; Supplemental Figs. S7 and S8).

Characterization of Flower-Specific TPSs

VOC-gene correlation analysis suggested that two flower-specific g-subfamily TPSs (Supplemental Fig. S6; CsaTPS03 encoded by *Csa1G068570* and CsaTPS01 encoded by *Csa1G066550*) were probably involved in linalool production in the male and female flowers of cucumber. We were unable to obtain the open reading frame of *CsaTPS03* due to its low level of transcription. The open reading frame of *CsaTPS01* encodes a peptide of 598 amino acids. TPS assays using GPP or FPP as the substrate clearly demonstrated that CsaTPS01 also was a bifunctional TPS that produced linalool from GPP and trans-nerolidol from FPP (Fig. 7; Supplemental Fig. S7). The linalool synthase activity, combined with the plastid localization (Fig. 6), suggested that CsaTPS01 was probably responsible for linalool emission from both male flowers and female flowers in cucumber plants.

Characterization of Fruit-Specific TPSs

According to our VOC-gene correlation analysis, CsaTPS15 (encoded by *Csa3G040850*) was predicted to play a role in sesquiterpene production in the fruits of cucumber (Table I). The open reading frame of *CsaTPS15* encodes a peptide of 568 amino acids that, based on phylogenetic analysis, belongs to the TPS-a subfamily. Although CsaTPS15 and CsaTPS12 share only 57% amino acid identity, enzymatic assays showed that CsaTPS15 produced similar terpenes to CsaTPS12, regardless of which prenyl pyrophosphate, GPP or FPP, was used as the assayed substrate. All sesquiterpenes (cis-α-bergamotene, trans-β-farnesene, β-bisabolene, aromadendrene, α-patchoulene, and α-bisabolene) detected in cucumber fruits were produced by CsaTPS15 when using FPP as a substrate (Fig. 7; Supplemental Fig. S8).

Heteromeric G(G)PPS Provides GPP for Monoterpene Production in Cucumber

It is known that the heteromeric G(G)PPS enzyme complexes involved in monoterpene biosynthesis are composed of a large subunit (LSU; usually a functional GGPPS) and a small subunit (SSU; type I and type II; Burke et al., 1999; Orlova et al., 2009; Wang and Dixon, 2009; Gutensohn et al., 2013; Chen et al., 2015). The cucumber genome contains both type I and type II SSU genes (CsaSSU I encoded by *Csa7G211090* and CsaSSU II encoded by *Csa7G017680*). Both CsaSSU I and CsaSSU II lost SARM domains (second Asp-rich domain; DDxxD, where x is any amino acid) and have two conserved CxxxC motifs (where x is any amino acid; Fig. 8A). Our VOC-gene correlation analysis suggested that *CsaSSU I* was probably involved in monoterpene biosynthesis in cucumber (Table I).

Given that there are two GGPPS-like genes and two *SSU* genes in the cucumber genome (Supplemental Data Set S7), we performed yeast two-hybrid assays to evaluate the interactions among these proteins. Both the yeast

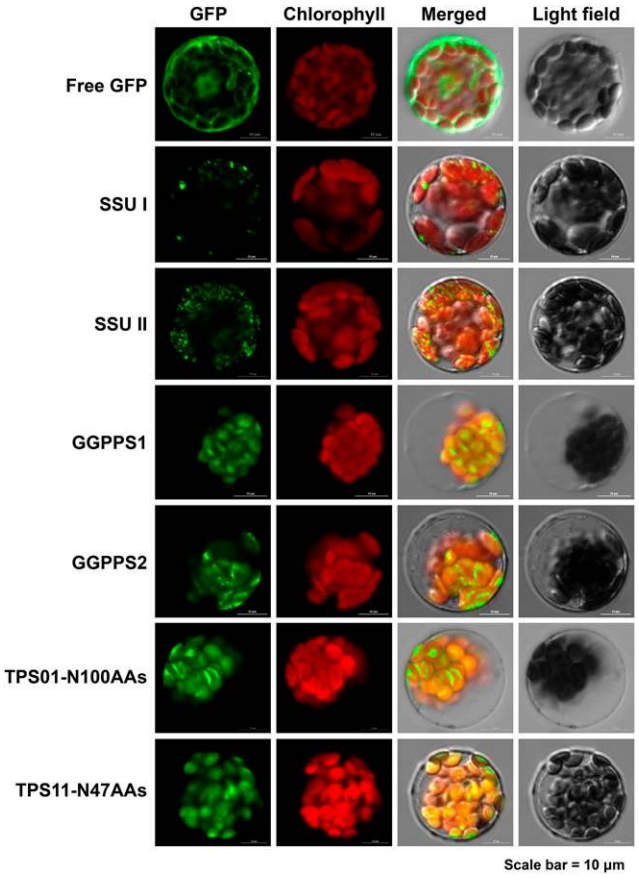

Figure 6. Subcellular localization of free GFP, four IDS-GFPs, TPS01-GFP, and TPS11-GFP in Arabidopsis leaf protoplasts. Chloroplasts have red chlorophyll autofluorescence. A total of 100 N-terminal amino acids of TPS01 and 47 N-terminal amino acids of TPS11 were fused with GFP in this experiment.

two-hybrid results and immunoblotting experiments with specific antibodies of His tag and S tag in *E. coli* showed that both of the GGPPS-like proteins associated with SSU I or SSU II to form heteromeric protein complexes (Fig. 9, A and B).

The N-terminally truncated proteins of the putative heteromeric G(G)PP-LSU and G(G)PP-SSU complexes were subcloned into the pRSFDuet-1 expression system, and these proteins were produced in *E. coli* with His-tagged LSUs and S-tagged SSUs (see "Materials and Methods"). The purified His-tagged heteromeric G(G)PP showed predominantly GGPPS activity when using DMAPP and [1-^{14}C]IPP as substrates, although the GGPPS1/SSU I complex showed mostly GPPS activity (greater than 70% of the product was GPP). A negligible amount of GPP was produced in the heterodimeric CsaGGPPS1/CsaSSU II and CsaGGPPS2/CsaSSU II assays, inconsistent with the results observed for the Arabidopsis (*Arabidopsis thaliana*) LSU/SSU II complex (AtSSU II was encoded by *At4g38460* and LSU was encoded by *At4g36810*). The Arabidopsis LSU/SSU II complex produced GPP at a higher efficiency than LSU itself (Wang and Dixon, 2009; Chen et al., 2015). Additionally, in all of the heteromeric CsaGGPPS/CsaSSU assays, GGPP was the main product produced when GPP or FPP was used as cosubstrate (Fig. 9C). Thus, we conclude that heteromeric CsaGGPPS1 and CsaSSU I provide GPP for monoterpene production in cucumber.

Kinetic characterization revealed that the apparent K_m value for DMAPP (27.7 μM) of the CsaGGPPS1/CsaSSU I complex was around 3-fold lower than that of CsaGGPPS1 alone (9.5 μM). Conversely, SSU I increased the affinity of CsaGGPPS2 for DMAPP about 2-fold (from 28.3 to 11.9 μM; Table II). As shown in Table II, the results clearly show that binding with CsaSSU II did not alter the affinity for DMAPP of either CsaGGPPS1 or CsaGGPPS2. Meanwhile, both CsaSSU I and CsaSSU II increased the enzymatic efficiency of CsaGGPPS by 2- to 4-fold (Table II).

DISCUSSION

It is well documented that the plant volatile chemicals, together with the nonvolatile specialized chemicals, play vital roles in both direct and indirect adaptions of plants to their environments (Holopainen and Gershenzon, 2010; Dudareva et al., 2013). Although the plant VOCs are vital elements of a plant's chemotype, research into plant VOC biosynthesis has largely lagged behind the pace of research into nonvolatile specialized chemicals in plants. Large-scale omics techniques had been applied successfully for the elucidation of a metabolic network of nonvolatile specialized chemicals in Arabidopsis (Matsuda et al., 2010; Saito and Matsuda, 2010). In this study, we demonstrate that an integrative analysis of nontargeted volatile profiling and transcriptome data can be applied for rapid elucidations of VOC biosynthetic pathways in a plant species of interest. VOC sampling and analysis play critical roles for reliable VOC-gene correlation analysis. Tholl et al. (2006) had comprehensively reviewed various SPME approaches (static headspace and dynamic headspace) for plant VOC collection and analysis. In this study, the cucumber plant VOCs were profiled in a nontargeted manner using static headspace SPME-GC-MS, after preparation in salt solution (Tholl et al., 2006; Gonda et al., 2010). Our results demonstrated that the analytical method used in this study has a good reproducibility and can be easily applied to other plant species to generate a high-resolution atlas of volatile chemicals. A total of 85 volatile compounds were detected from cucumber plants; however, more than two-thirds (61 of 85) of cucumber VOCs remain structurally uncharacterized. To get more structural information on cucumber VOCs, high-resolution mass spectrometry, such as quadrupole time-of-flight mass spectrometry, should be applied (Kusano et al., 2013). In addition to the VOC profiling, additional metabolomics data generated from liquid chromatography-mass spectrometry analysis will, in the future, provide more metabolic

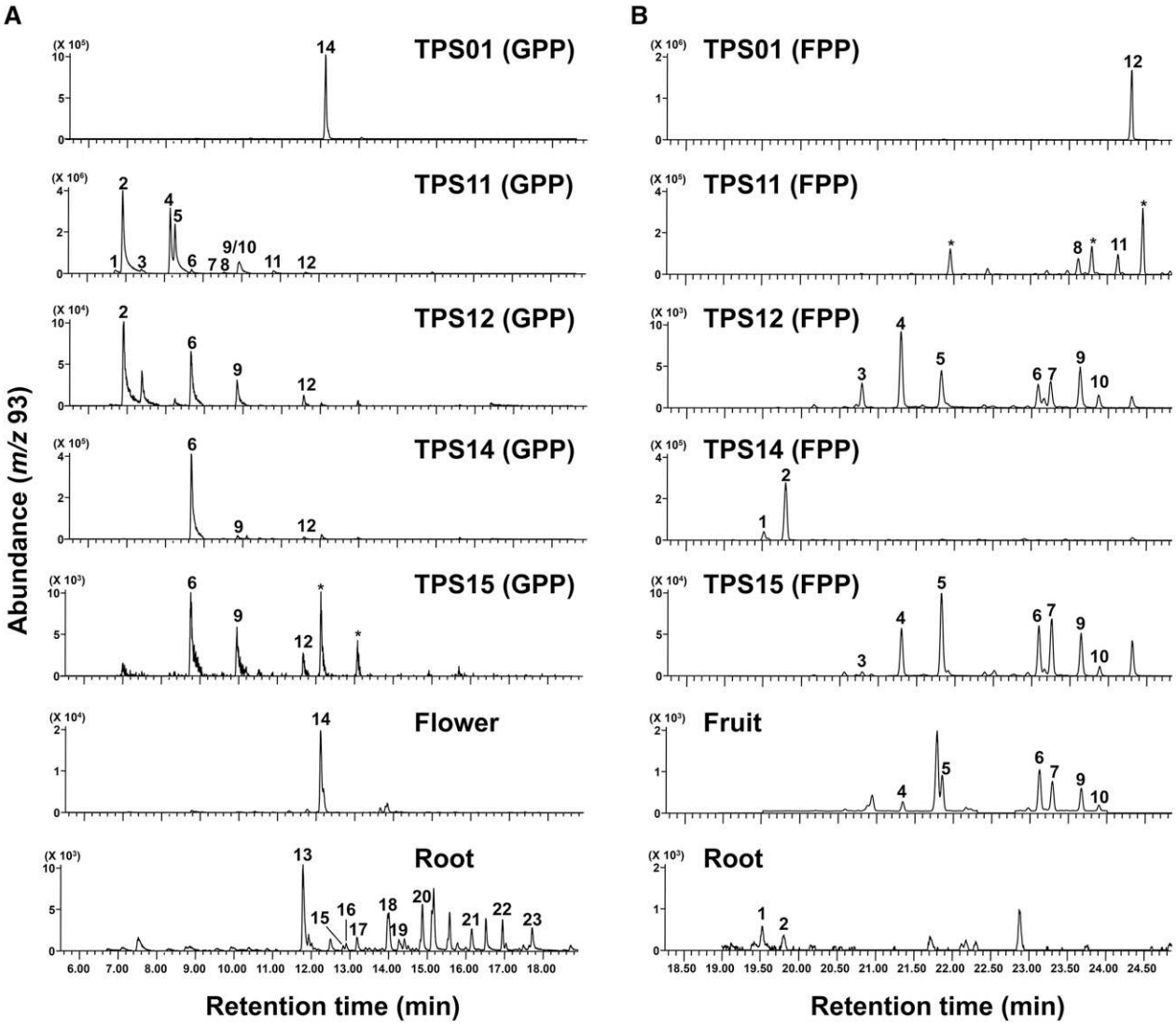

Figure 7. In vitro enzymatic assays of the CsaTPSs. The gas chromatographs (selected ion monitoring [SIM] mode, mass-to-charge ratio [*m/z*] of 93) of volatile terpene-producing tissues (monoterpenes produced from flower and root tissues; sesquiterpenes produced from fruit and root tissues) were used for comparison. Peaks marked with asterisks were tentatively identified as monoterpene or sesquiterpene, the mass spectra of which have no hit in the NIST 08 library. A, Representative gas chromatographs (SIM mode, *m/z* 93) of product profiling of CsaTPSs using GPP as a substrate (from top to bottom, TPS01, TPS11, TPS12, TPS14, and TPS15). Peaks (ordered by retention time) are as follows: 1, β-thujene; 2, α-pinene; 3, camphene; 4, sabinene; 5, β-pinene; 6, β-myrcene; 7, α-phellandrene; 8, α-terpinen; 9, limonene; 10, eucalyptol; 11, γ-terpinen; 12, *p*-mentha-1,4(8)-diene; 13, α-campholenal; 14, linalool; 15, limonene oxide; 16, (*E*)-limonene oxide; 17, pinocarveol; 18, (*Z*)-limonene oxide; 19, ipsdienol; 20, myrtenol; 21, (*Z*)-3,7-dimethyl-2,6-octadienal; 22, (*E*)-3,7-dimethyl-2,6-octadienal; 23, perilla alcohol. Among them, peaks 1, 4, 6 to 8, 11 to 13, 15, 16, 18, 19, 21, and 22 were tentatively identified as monoterpenes based on spectral matching to spectra in the NIST 08 library; and peaks 2, 3, 5, 9, 10, 14, 17, 20, and 23 were determined by comparing with monoterpene standards (for detailed mass spectra, see Supplemental Fig. S7). B, Representative gas chromatographs (SIM mode, *m/z* 93) of product profiling of CsaTPSs using FPP as a substrate (from top to bottom, TPS01, TPS11, TPS12, TPS14, and TPS15). Peaks (ordered by retention time) are as follows: 1, cyclosativene; 2, α-cubebene; 3, trans-α-bergamotene; 4, cis-α-bergamotene; 5, trans-β-farnesene; 6, β-bisabolene; 7, aromadendrene; 8, α-himachalene; 9, α-patchoulene; 10, α-bisabolene; 11, β-sesquiphellandrene; 12, trans-nerolidol. Among them, peaks 2 to 11 were tentatively identified as sesquiterpenes based on spectral matching to spectra in the NIST 08 library, and peaks 1 and 12 were determined by comparing with sesquiterpene standards (for detailed mass spectra, see Supplemental Fig. S8).

information that we plan to use to complete our exploration and characterization of the metabolic network in cucumber.

Although the cucumber VOC data from this study do not reflect the natural emission of VOCs of intact organs, which is limited for the dissection of ecological functions

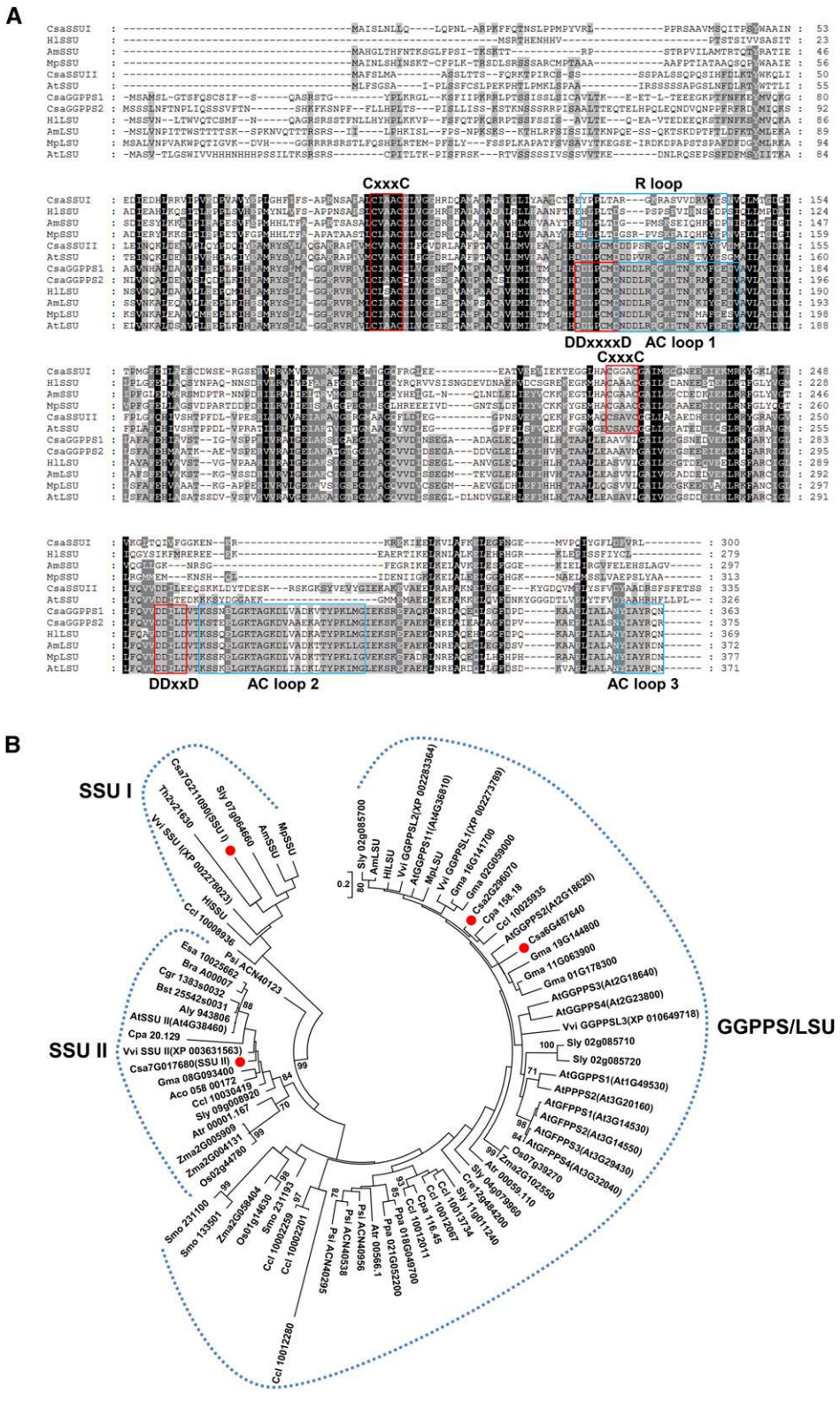

Figure 8. Sequence and evolutionary analysis of plant G(G)PPS proteins. A, Alignment of cucumber G(G)PPS amino acid sequences with functionally characterized SSUs and LSUs from other species. Two conserved CxxxC motifs for the SSU proteins and two conserved Asp-rich motifs for the LSU proteins are boxed in red. The R-loop of SSU and the AC-loops of LSU are boxed in blue. GenBank accession numbers are as follows: AmLSU, AAS82860; AmSSU, AAS82859; AtLSU, NP_195399; AtSSU, NP_195558; HlLSU, ACQ90682; HlSSU, ACQ90681; MpLSU, AAF08793; MpSSU, AF182827. B, Phylogenetic analysis of SSU and LSU of heterodimeric G(G)PPSs. Bootstrap values (based on 1,000 replications) greater than 70% are shown for the corresponding nodes. The experimentally confirmed cucumber proteins are indicated with red circles. The scale measures evolutionary distance in substitutions per site. Species abbreviations are as follows: Aco, *Aquilegia coerulea*; Aly, *Arabidopsis lyrata*; Am, *Antirrhinum majus*; At, *Arabidopsis thaliana*; Atr, *Amborella trichopoda*; Bra, *Brassica rapa*; Bst, *Boechera stricta*; Ccl, *Citrus clementine*; Cgr, *Capsella grandiflora*; Cpa, *Carica papaya*; Cre, *Chlamydomonas reinhardtii*; Cru, *Capsella rubella*; Esa, *Eutrema salsugineum*; Hl, *Humulus lupulus*; Gma, *Glycine max*; Mp, *Mentha* × *piperita*; Os, *Oryza sativa*; Psi, *Picea sitchensis*; Ppa, *Physcomitrella patens*; Sly, *Solanum lycopersicum*; Smo, *Selaginella moellendorffii*; Tha, *Tarenaya hassleriana*; Vvi, *Vitis vinifera*; Zma, *Zea mays*.

of cucumber VOCs, the integration of VOC profiling and transcriptome data is a feasible strategy for rapid VOC pathway elucidation at the molecular level. As shown in Table I, the candidate genes for the main known VOCs (terpenoids, phenylpropanoids/benzenoids, and C6/C9 alcohols/aldehydes) were determined based on known biochemical knowledge and correlation analysis. As a case study, we functionally characterized the terpene

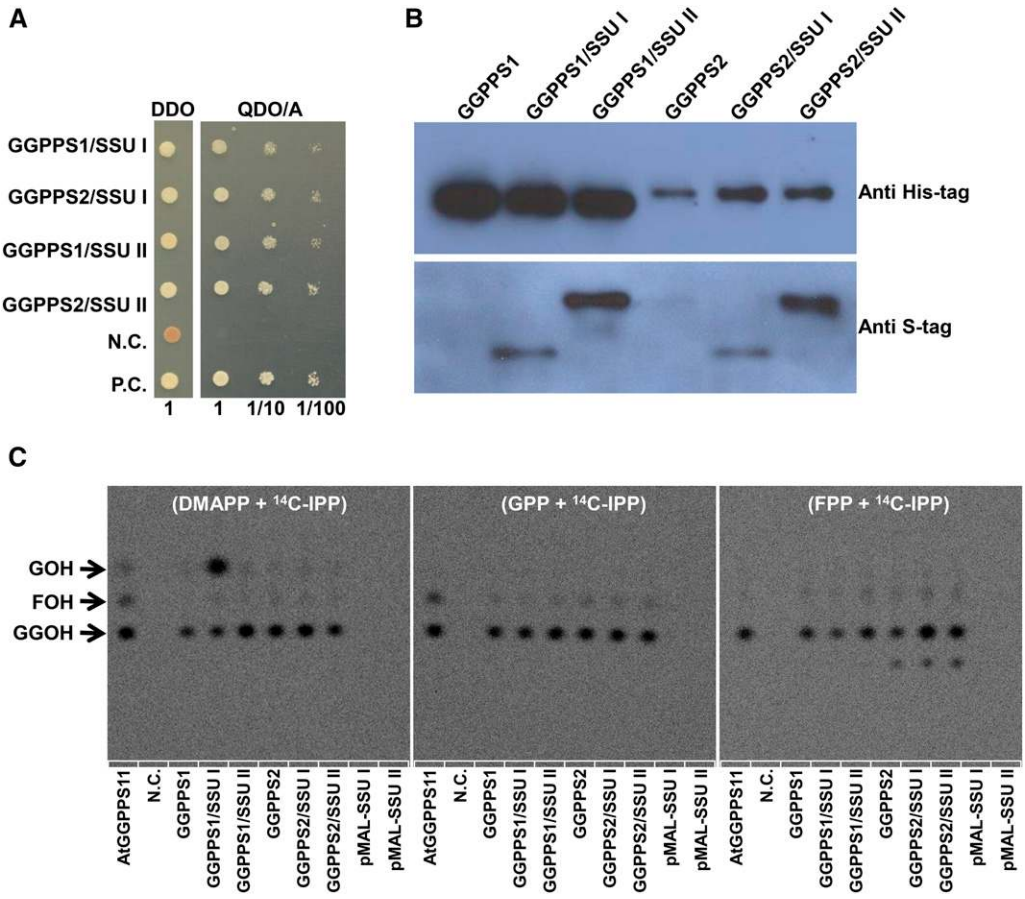

Figure 9. Heteromeric geranyl(geranyl) diphosphate synthase, composed of CsaGGPPS and SSU, is involved in monoterpene biosynthesis in cucumber. A, Protein-protein interactions between the CsaGGPPS1/2 and CsaSSU I/II proteins in a yeast two-hybrid system. Synthetic dextrose (double dropout [DDO]; −Leu, −Trp) was used to select the bait and prey constructs; synthetic dextrose (quadruple dropout [QDO]/A; −Leu, −Trp, −His, −adenine, + aureobasidin A) was used to select the interactions between the GGPPS proteins. N.C., Negative control (pGBKT7-Lam and pGADT7-T constructs); P.C., positive control. The predicted N-terminal signal peptide was removed for all of the GGPPS and SSU proteins tested in this experiment. B, Protein complexes (1 μg of protein) purified with nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography (His tagged) were separated by SDS-PAGE, transferred onto nitrocellulose membranes, and blotted with anti-His tag (for GGPPS detection) and anti-S tag (for SSU detection) monoclonal antibodies. N-terminally His-tagged G(G)PPS1/2 (calculated mass of 36,198/37.005 kD) and C-terminally S-tagged G(G)PPS.SSU I/II (calculated mass of 31,487/38.071 kD) were used. C, In vitro G(G)PPS assay. The purified recombinant heteromeric G(G)PPSs were assayed using [^{14}C]IPP and different allylic substrate (DMAPP, GPP, or FPP). The reaction products were separated via reverse-phase thin-layer chromatography (TLC), and the positions of prenols with various chain lengths are indicated with arrows: GOH, geranol; FOH, farnesol; GGOH, geranylgeranol. AtGGPPS11 enzyme was used as a positive control, and boiled AtGGPPS11 enzyme was used as a negative control in this assay.

biosynthesis pathway in cucumber, with a focus on the TPS and IDS genes selected by VOC-gene correlation analysis. In vitro biochemical assays, together with expression pattern and subcellular localization information, suggested that TPS11/TPS14, TPS01, and TPS15 were responsible for volatile terpenoid production in the roots, flowers, and fruit tissues of cucumber plants, respectively. The results also showed that all tested cucumber TPSs are bifunctional enzymes in vitro: they catalyze GPP to form monoterpenes and FPP to form sesquiterpenes at similar efficiency (Fig. 7). The previously reported farnesene synthase (Sester-TPS) from cucumber (CsaTPS19 in this study), selected by a VOC-gene coresponse approach, also catalyzed the GPP to (*E*)-β-ocimene (Mercke et al., 2004). Both the enzymatic promiscuity (bifunctional) of tested TPSs and the tandem localization of TPS subfamily genes on cucumber chromosomes suggested that the cucumber TPS gene family is still undergoing dramatic evolution. These characteristics of the TPS gene family also increase the possibility of the subfunctionalization and neofunctionalization of TPSs during cucumber evolution (Pichersky and Gang, 2000; Weng, 2014).

Plant root systems, which interact with a huge diversity of soil organisms (bacteria and fungi), are able to produce and release VOCs, and possible ecological

Table II. *Kinetic parameters (toward DMAPP) for recombinant heteromeric and heteromeric G(G)PPS of cucumber*

All results are presented as means ± SD of triplicate experiments.

Enzyme	K_m [a]	K_{cat}/K_m
	μM	s^{-1} M^{-1}
CsaGGPPS1	9.5 ± 0.1	52.0 ± 2.4
CsaGGPPS1/CsaSSU I	21.7 ± 2.5	173.1 ± 12.3
CsaGGPPS1/CsaSSU II	8.2 ± 0.1	133.1 ± 20.1
CsaGGPS2	28.3 ± 5.5	60.8 ± 11.9
CsaGGPS2/CsaSSU I	11.9 ± 5.4	251.9 ± 37.7
CsaGGPS2/CsaSSU II	26.7 ± 1.6	122.9 ± 4.1

[a]With 30 μM IPP as cosubstrate.

functions of root-emitted VOCs have been proposed (Peñuelas et al., 2014, and refs. therein). In light of the tissue specificity and enzyme assay results of our study, the in planta functions of terpenoids from cucumber mature roots merit further research attention. One interesting discovery from our study is that most monoterpenes produced by CsaTPS11 are the direct precursors of the monoterpene oxide compounds that were detected in mature root tissues of cucumber (Supplemental Fig. S1). VOC-gene correlation analysis also gave us some clues that were useful in selecting several *P450* genes (there are 188 putatively annotated *P450* genes in the cucumber genome; Table I; Supplemental Data Set S4) that seem likely to participate in the oxygenation process of monoterpenes in mature roots of cucumber plants (Ginglinger et al., 2013; Weitzel and Simonsen, 2015). The functional characterization of several of these candidate *P450* genes is currently being conducted in our laboratory.

It is noteworthy that all tested TPSs in this study were found to be bifunctional enzymes; that is, these enzymes were able to utilize both GPP (derived from the plastid MEP pathway) and FPP (derived from the cytosolic MVA pathway) as substrates. The expression pattern and subcellular localization information further supported the in planta functions of these TPSs. However, the exact functions and biological significance of these *TPS* genes in cucumber will need to be elucidated by gain- or loss-of-function approaches like those used in model plants such as Arabidopsis (Tholl et al., 2005). Recently reported transformation protocols for cucumber should facilitate the investigation of the enzymatic hypotheses generated from this study (Nanasato et al., 2013; Wang et al., 2015).

It has been documented that GPPS enzymes were evolved from GGPPS enzymes in gymnosperm plants (Burke and Croteau, 2002; Schmidt and Gershenzon, 2008). However, the angiosperm plants acquired two types of SSU proteins to form heteromeric G(G)PPS by binding a GGPPS-like protein (Burke et al., 1999; Wang and Dixon, 2009; Chen et al., 2015). Our phylogenetic analysis indicated that the type I and type II SSUs share a common ancestor that was probably a gymnosperm G(G)PPS (Fig. 9; Wang and Dixon, 2009; Coman et al., 2014). As shown in the phylogenetic tree, type II *SSU* genes have been found in the genomes of all sequenced angiosperm plants. However, type I *SSU* genes occur only sporadically in angiosperm plants, such as in hop (*Humulus lupulus*), peppermint (*Mentha* × *piperita*), cucumber (this study), and others (Fig. 8B). One plausible explanation for this is that most *SSU I* genes may have been lost during the evolution of angiosperm plants, and only those that conferred a selective advantage (increasing the pollination efficiency or biotic resistance, etc.) to a host plant were retained in a particular genome. Given the fact that most cucumbers are monoecious (i.e. having both male and female flowers on a single plant), the flower-specific compound linalool (in both male and female flowers), which is produced by TPS01 and CsaSSU1 from the MEP pathway, may play a key role for attracting pollinators (Dudareva et al., 2013; Zhang et al., 2015).

It is of interest that the expression of the *SSU I* genes that were retained in plant genomes is restricted to nongreen tissues, such as flowers, roots, and glandular trichomes (Burke et al., 1999; Tholl et al., 2004; Wang and Dixon, 2009). It is possible that the elimination of GGPPS activity (LSU) by binding with SSU I will affect chlorophyll biosynthesis, or the formation of other important terpenoids, in green tissues. Consistent with this hypothesis, overexpression of a snapdragon (*Antirrhinum majus*) *SSU I* gene driven by the 35S promoter in tobacco (*Nicotiana tabacum*) resulted in leaf chlorosis and dwarfism (Orlova et al., 2009). However, tissue-specific overexpression of a snapdragon *SSU I* gene driven by the fruit ripening-specific *PG* promoter in tomato fruits (a nongreen tissue) did not cause any developmental defects; however, this did alter the metabolic flux among compounds in terpenoid metabolism (Gutensohn et al., 2013). Interestingly, the cucumber SSU I protein increased the GPP provision in two tissues of a single plant, roots and flowers (both are nongreen tissues), which was not described previously. It should be noted that the expression of *SSU I* in mature root tissue was not checked in previous SSU-related studies (Burke et al., 1999; Tholl et al., 2004; Wang and Dixon, 2009). Our study provides a more comprehensive VOC and gene atlas of a plant species than ever before.

In addition to the terpenoid biosynthetic pathways prioritized in this study, the information generated in our VOC-gene correlation analysis will be of use for future efforts to elucidate the biosynthetic pathways of other cucumber VOCs (Table I). Our next efforts will target the fresh green note flavor (C9 aldehydes) and other VOCs in cucumber fruits. Although cucumber fruits produce the fresh green note flavor extensively, knowledge about which VOCs contribute to flavors that match consumer preferences remains lacking. Recently, 115 cucumber cultivars, covering four geographic groups (30 lines forming an Indian group [the wild cucumbers], 37 lines from East Asia, 19 landraces from Xishuangbanna in southwest China, and 29 lines as a Eurasian group), were resequenced (Qi et al., 2013). The VOC profiling data of these 115 cucumber lines, combined with sensory analysis by a consumer panel, will provide insight into the genetic basis of cucumber fruit flavor quality and valuable information for the flavor improvement of cucumber fruit (Tieman et al., 2012). Elucidation of the biosynthetic

pathway(s) of consumer-preferred VOCs will be facilitated through the type of VOC-gene correlation analysis that we have demonstrated in this study.

MATERIALS AND METHODS

Plant Materials and Chemicals

The cucumber plants (*Cucumis sativus* '9930') used in this study were grown in the experimental station of the Institute of Genetics and Developmental Biology of the Chinese Academy of Sciences in Beijing. To obtain the sample tissues from 4-week-old seedlings (root, stem, cotyledon, and true leaf), seeds were sown on one-half-strength Murashige and Skoog medium after sterilization and grown in a growth incubator (28°C/22°C and 12-h/12-h light/dark photoperiod). The other 19 sample tissues, including root, stem, young leaf (the second node from the top), old leaf (the fifth node from the top), flower, cotyledon, hypocotyl, and fruit at different developmental stages, were collected from 12-week-old cucumber plants. For each biological replicate of sample collection, the tissue was collected from at least four individual plants; this tissue was combined, and the mixed material was quickly ground and frozen using liquid nitrogen and then divided into two aliquots: one aliquot of frozen samples (around 100 mg) was subsequently used for VOC analysis; another aliquot (around 100 mg) was used for RNA preparation and sequencing after being combined with other collections of the same tissue. Two to four biological replicates, depending on the availability of plant materials, were carried out for the VOC analysis of each tissue. The combined plant materials, from two to four independent collections, were used for RNA isolation and RNA sequencing.

All chemical standards used in this study were purchased from Sigma-Aldrich, with the exception that the radiolabeled [1-^{14}C]IPP (55 mCi mmol^{-1}) was purchased from American Radiolabeled Chemicals. All DNA constructs generated in this study were verified by sequencing.

VOC Sampling and GC-MS Analysis

The cucumber sample powder was sealed in a 50-mL tube and stored at −80°C for no longer than 2 weeks before extraction and analysis. A total of 100 ± 0.5 mg of tissue powder was weighed and transferred into a 4-mL glass vial (Agilent Technologies) filled with 400 μL of 20% NaCl solution (used to disrupt the activity of endogenous enzymes). The volatile compound 2-heptone was added as an internal standard to a final concentration of 0.125 ng μL^{-1}. For volatile collection, a 100-μm fused silica fiber coated with DVB/CAR/PDMS (Sigma-Aldrich) was used (Kusano et al., 2013; Risticevic and Pawliszyn, 2013). A cleaned fiber was inserted into the vial (preheated at 30°C for 5 min) and exposed to the headspace for 30 min at 30°C. The SPME fiber was introduced into the injector port of a GC-MS instrument (Agilent 7890A GC-5975C MSD) equipped with a DB-5 mass spectrometry column (30 m × 0.25 mm × 0.25 μm) and held for 30 s. The injector (splitless mode) temperature was set at 250°C. The oven temperature was initially set at 40°C, held for 1.5 min, and then increased to 150°C at a rate of 5°C min^{-1}. Subsequently, the temperature was ramped at 15°C min^{-1} to 260°C and then held for 10 min. The temperature of the quadruple mass analyzer was set at 150°C. Mass spectra were acquired within a scanning range of 50 to 400 m/z. Metabolites were identified by comparing spectra and retention times with commercially available standards or tentatively identified by spectral matching with entries in the NIST 08 library (NIST and Wiley libraries). Two to four independent (depending on the available amount of plant material) biological replicates were analyzed in this experiment. The peak area of each metabolite, acquired in SIM mode (signal-noise > 5), was normalized to the peak area of 2-heptone prior to further data processing.

RNAseq and Transcriptome Analyses

Total RNA was extracted as described previously (Wang et al., 2008). The isolated RNA, from 23 combined cucumber tissues, was used for cDNA library construction only after it passed a quality evaluation. The cDNA libraries were sequenced on a HiSeq 2000 analyzer in a paired-end (2 × 100 bp) and indexed sequencing mode. The cucumber genome and annotation (version 2) were downloaded from http://www.icugi.org/cgi-bin/ICuGI/index.cgi. RNAseq analysis for the 23 sampled cucumber tissues was performed based on the protocol detailed by Trapnell et al. (2012). First, we built a reference index by Bowtie-build2 (version 2.0.5). Then, TopHat (version 2.0.8) was used to align the cleaned raw reads to the reference index. Unique mapped reads were filtered with a mapping quality threshold over 15 ($P < 0.05$). Cufflinks (version 2.2.1) was used to assemble the transcripts according to the cucumber annotations. The expression quantity of cucumber unigenes is expressed in terms of FPKM. Default parameter settings were used for these programs. The threshold of detectable expression was set to 0.5 FPKM. Based on Blast2GO annotation, the overall distribution frequency of the Gene Ontology categories in the cucumber genome was obtained.

qRT-PCR Analysis

qRT-PCR analysis, with SYBR Premix Ex Taq II (Tli RNaseH Plus; TaKaRa), was performed on the CFX96 real-time system (Bio-Rad). The gene encoding the ubiquitin extension protein (GenBank accession no. AY372537) was chosen for use as the internal standard gene for the qRT-PCR experiments (Wan et al., 2010), and the comparative C_T method was applied to calculate gene relative expression level (Schmittgen and Livak, 2008). Every measurement was performed with at least three biological replicates. The primers used for qRT-PCR analysis are listed in Supplemental Table S1.

Statistical Analysis

The expression data and metabolite data were standardized as *Z* scores. These data were then used for PCA and the calculation of PCCs. PCCs were calculated both among the gene expression data and between gene expression and metabolite accumulation data. PCA was conducted using the princomp R package, and the PCA results were visualized using the scatterplot3d R package. Clustering analysis of the 85 VOCs in the 23 tissues was performed using Cluster 3.0 (de Hoon et al., 2004). Distance measures were specified by Spearman correlation coefficient values. The correlation network was visualized by means of Cytoscape 3.2.1 software (Shannon et al., 2003). *H* values were calculated using a previously described procedure (Schug et al., 2005). $H \leq 2$ was used to screen for tissue-specific genes.

Generation, Expression, and Purification of Recombinant CsaTPSs and CsaIDSs

The protein expression constructs of cucumber IDS and TPS genes were generated following a previously described strategy with minor modifications (Wang et al., 2008; Wang and Dixon, 2009). The truncated *CsaGGPPS1* and *CsaGGPPS2* cDNAs were cloned into the expression vector pRSFDuet-1 (Novagen), while the truncated *CsaTPS01*, *CsaTPS11*, *CsaTPS14*, *CsaTPS15*, *CsaSSU I*, and *CsaSSU II* cDNAs were cloned into pMAL-c2x vector (New England Biolabs). All constructs were transformed into *Escherichia coli* strain BL21 (DE3) for protein production. The resulting His-tagged GGPPSs and MBP-tagged proteins (TPSs and SSUs) were purified using Ni-NTA (for His-tagged proteins) and amylose resin (for MBP-tagged proteins) affinity chromatography, respectively. For coexpression of *CsaGGPPS* and *SSU* in *E. coli*, the truncated cDNAs were cloned into different cassettes of the pRSFDuet-1 vector (Novagen), resulting in His-tagged GGPPS and S-tagged SSU. The GGPPS/SSU complexes were purified using Ni-NTA affinity chromatography. Quantification and purity evaluation of the recombinant proteins were performed using SDS-PAGE with BSA as a standard. The primers used in this experiment are detailed in Supplemental Table S1.

TPS Assays

For the TPS enzyme assays, the incubation mixture (final volume, 500 μL) contained 50 mM HEPES (pH 7.5), 20 mM $MgCl_2$, 5 mM DTT, and 5 μg mL^{-1} various prenyl pyrophosphates (GPP, FPP, or GGPP) and was set into a 4-mL glass vial (Agilent Technologies). The reactions were initiated by the addition of 1 to 5 μg of purified recombinant TPS protein. The assays were incubated for 60 min at 30°C. For volatile product collection, a 100-μm fused silica fiber coated with DVB/CAR/PDMS was applied and exposed to the headspace of the reaction mixture for 30 min at 30°C. The captured volatile products by SPME were analyzed by GC-MS as described above. Assays with heat-denatured protein were used as controls.

IDS Assays and Enzyme Characterization

The IDS assays were performed as described previously (Wang and Dixon, 2009). Briefly, 100 μL of reaction mixture, containing 250 mM MOPS (pH 7), 2 mM dithiothreitol, 10 mM $MgCl_2$, 1 μL of prenyl pyrophosphate (DMAPP,

GPP, or FPP), 1 μL of [1-^{14}C]IPP, and 10 μg of purified recombinant protein, was incubated for 60 min at 30°C. The resulting prenyl alcohols, released by bovine intestine alkaline phosphatase and potato apyrase (Sigma-Aldrich), were separated using a reverse-phase (C_{18} silica gel-60 matrix, F254S) TLC plate (Merck). The TLC plate was developed with acetone:water (9:1, v/v) solvent. The radiolabeled spots on the TLC plate were measured by phosphorimaging (Typhoon trio) and quantified with ImageJ software. A DMAPP concentration series of 2, 5, 10, 20, and 50 μM, together with 30 μM IPP (cold IPP:[1-^{14}C]IPP = 3.3:1), was used in the DMAPP kinetics experiments. The apparent K_m and K_{cat} data were calculated using Hanes plots (Hyper32, version 1.0.0).

Yeast Two-Hybrid Assays

Protein-protein interactions between CsaGGPPSs and CsaSSUs were evaluated using the Matchmaker Gold Yeast Two-Hybrid System Kit, following the manufacturer's instructions (Clontech). The truncated *SSUs* were inserted into pGBKT7 vector, and the truncated *GGPPSs* were inserted into pGADT7 vectors. The vectors were then transformed into yeast strain Y2HGold for the testing of reciprocal interactions. The primers used in this experiment are listed in Supplemental Table S1.

Coimmunoprecipitation Assays of CsaGGPPS and CsaSSU

N-terminal His-tagged *CsaGGPPS* and C-terminal S-tagged *CasSSU* were cloned into different cassettes of pRSFDuet-1 vector (Novagen). The correct plasmid, confirmed by sequencing, was introduced into *E. coli* BL21 (DE3) for protein production. The GGPPS/SSU complexes, after purification using Ni-NTA affinity chromatography, were resuspended in SDS-PAGE sample buffer. The protein samples (1 μg of protein) were separated by 12% SDS-PAGE, transferred onto a polyvinylidene difluoride membrane, and blotted with anti-His or anti-S antibody.

Subcellular Localization of CsaG(G)PPSs and CsaTPSs

Constructs for full-length CsaG(G)PPSs and signal peptide of CsaTPS01/11-GFP fusion proteins (pJIT163-hGFP vector), Arabidopsis (*Arabidopsis thaliana*) leaf protoplast preparation, transformation, and image collection using laser scanning confocal microscopy were performed as described previously (Yoo et al., 2007; Xu et al., 2013).

Sequence Alignment and Phylogenetic Analysis

Most plant GGPPS (or LSU) and SSU sequences were extracted from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) and Phytozome 10.0 (http://www.phytozome.net) databases (for detailed sequence information, see Supplemental Data Set S9). Amino acid sequence alignment of plant GGPPS (or LSU) and SSU was conducted with ClustalW, and the final file was generated using GeneDoc (http://www.psc.edu/index.php/user-resources/software/genedoc). A maximum likelihood tree, with 1,000 replicates for bootstrap values, was constructed using MEGA6 software (Tamura et al., 2013).

Accession Numbers

RNAseq raw sequence data of 23 cucumber tissues generated from this study can be found in the National Center for Biotechnology Information Short Read Archive database under the following accession number: SRA071224.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Proposed monoterpenoid biosynthetic pathway in cucumber root.

Supplemental Figure S2. Proposed sesquiterpenoid biosynthetic pathway in cucumber.

Supplemental Figure S3. Overall statistics of RNAseq data.

Supplemental Figure S4. Gene Ontology categories of the 21,788 expressed unigenes from cucumber.

Supplemental Figure S5. Correlation analysis of cucumber catalytic genes and VOC profiling data.

Supplemental Figure S6. Genome distribution of 24 cucumber TPS genes and phylogenetic tree of the corresponding TPS proteins.

Supplemental Figure S7. Mass spectra of the monoterpenes described in Figure 7A.

Supplemental Figure S8. Mass spectra of the sesquiterpenes described in Figure 7B.

Supplemental Table S1. Primers used in this study.

Supplemental Data Set S1. Volatile profiling of 23 tissues from cucumber plants (normalized to the internal standard).

Supplemental Data Set S2. Volatile profiling (peak area) of 23 tissues from cucumber plants.

Supplemental Data Set S3. Tissue specificity of all cucumber genes, FPKM.

Supplemental Data Set S4. VOC-gene correlation analysis in cucumber.

Supplemental Data Set S5. Tissue-specific expression of the genes involved in the MVA pathway.

Supplemental Data Set S6. Tissue-specific expression of the genes involved in the MEP pathway.

Supplemental Data Set S7. Tissue-specific expression pattern of 24 cucumber TPSs (more than 200 amino acids in length).

Supplemental Data Set S8. Tissue-specific expression pattern of eight short-chain prenyltransferase genes from cucumber.

Supplemental Data Set S9. Amino acid sequences of plant GGPPS and SSU used for phylogenetic analysis.

Received July 5, 2016; accepted July 19, 2016; published July 25, 2016.

LITERATURE CITED

Burke C, Croteau R (2002) Geranyl diphosphate synthase from *Abies grandis*: cDNA isolation, functional expression, and characterization. Arch Biochem Biophys **405:** 130–136

Burke CC, Wildung MR, Croteau R (1999) Geranyl diphosphate synthase: cloning, expression, and characterization of this prenyltransferase as a heterodimer. Proc Natl Acad Sci USA **96:** 13062–13067

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) 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 **66:** 212–229

Chen Q, Fan D, Wang G (2015) Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in Arabidopsis flowers. Mol Plant **8:** 1434–1437

Coman D, Altenhoff A, Zoller S, Gruissem W, Vranová E (2014) Distinct evolutionary strategies in the GGPPS family from plants. Front Plant Sci **5:** 230

Degenhardt J, Köllner TG, Gershenzon J (2009) Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry **70:** 1621–1637

de Hoon MJL, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics **20:** 1453–1454

Dudareva N, Klempien A, Muhlemann JK, Kaplan I (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. New Phytol **198:** 16–32

Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I, Matsuba Y, Bonini ME, Schilmiller AL, Last RL, et al (2011) The tomato terpene synthase gene family. Plant Physiol **157:** 770–789

Ginglinger JF, Boachon B, Höfer R, Paetz C, Köllner TG, Miesch L, Lugan R, Baltenweck R, Mutterer J, Ullmann P, et al (2013) Gene coexpression analysis reveals complex metabolism of the monoterpene alcohol linalool in *Arabidopsis* flowers. Plant Cell **25:** 4640–4657

Gonda I, Bar E, Portnoy V, Lev S, Burger J, Schaffer AA, Tadmor Y, Gepstein S, Giovannoni JJ, Katzir N, et al (2010) Branched-chain and aroxmatic amino acid catabolism into aroma volatiles in *Cucumis melo* L. fruit. J Exp Bot **61:** 1111–1123

Gutensohn M, Orlova I, Nguyen TTH, Davidovich-Rikanati R, Ferruzzi MG, Sitrit Y, Lewinsohn E, Pichersky E, Dudareva N (2013) Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. Plant J **75:** 351–363

Hofberger JA, Ramirez AM, van den Bergh E, Zhu X, Bouwmeester HJ, Schuurink RC, Schranz ME (2015) Large-scale evolutionary analysis of genes and supergene clusters from terpenoid modular pathways provides insights into metabolic diversification in flowering plants. PLoS ONE **10:** e0128808

Holopainen JK, Gershenzon J (2010) Multiple stress factors and the emission of plant VOCs. Trends Plant Sci **15:** 176–184

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, et al (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet **41:** 1275–1281

Jia Q, Chen F (2016) Catalytic functions of the isoprenyl diphosphate synthase superfamily in plants: a growing repertoire. Mol Plant **9:** 189–191

Kappers IF, Hoogerbrugge H, Bouwmeester HJ, Dicke M (2011) Variation in herbivory-induced volatiles among cucumber (*Cucumis sativus* L.) varieties has consequences for the attraction of carnivorous natural enemies. J Chem Ecol **37:** 150–160

Klee HJ (2010) Improving the flavor of fresh fruits: genomics, biochemistry, and biotechnology. New Phytol **187:** 44–56

Knudsen JT, Eriksson R, Gershenzon J, Stahl B (2006) Diversity and distribution of floral scent. Bot Rev **72:** 1–120

Kusano M, Iizuka Y, Kobayashi M, Fukushima A, Saito K (2013) Development of a direct headspace collection method from Arabidopsis seedlings using HS-SPME-GC-TOF-MS analysis. Metabolites **3:** 223–242

Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. Plant Physiol **152:** 566–578

Mercke P, Kappers IF, Verstappen FWA, Vorst O, Dicke M, Bouwmeester HJ (2004) Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. Plant Physiol **135:** 2012–2024

Metzker ML (2010) Sequencing technologies: the next generation. Nat Rev Genet **11:** 31–46

Muhlemann JK, Klempien A, Dudareva N (2014) Floral volatiles: from biosynthesis to function. Plant Cell Environ **37:** 1936–1949

Nanasato Y, Konagaya KI, Okuzaki A, Tsuda M, Tabei Y (2013) Improvement of Agrobacterium-mediated transformation of cucumber (*Cucumis sativus* L.) by combination of vacuum infiltration and co-cultivation on filter paper wicks. Plant Biotechnol Rep **7:** 267–276

Orlova I, Nagegowda DA, Kish CM, Gutensohn M, Maeda H, Varbanova M, Fridman E, Yamaguchi S, Hanada A, Kamiya Y, et al (2009) The small subunit of snapdragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. Plant Cell **21:** 4002–4017

Peñuelas J, Asensio D, Tholl D, Wenke K, Rosenkranz M, Piechulla B, Schnitzler JP (2014) Biogenic volatile emissions from the soil. Plant Cell Environ **37:** 1866–1891

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci **5:** 439–445

Qi J, Liu X, Shen D, Miao H, Xie B, Li X, Zeng P, Wang S, Shang Y, Gu X, et al (2013) A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. Nat Genet **45:** 1510–1515

Risticevic S, Pawliszyn J (2013) Solid-phase microextraction in targeted and nontargeted analysis: displacement and desorption effects. Anal Chem **85:** 8987–8995

Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics: 'majority report by precogs.' Trends Plant Sci **13:** 36–43

Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. Annu Rev Plant Biol **61:** 463–489

Sallaud C, Rontein D, Onillon S, Jabès F, Duffé P, Giacalone C, Thoraval S, Escoffier C, Herbette G, Leonhardt N, et al (2009) A novel pathway for sesquiterpene biosynthesis from *Z,Z*-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. Plant Cell **21:** 301–317

Schilmiller AL, Schauvinhold I, Larson M, Xu R, Charbonneau AL, Schmidt A, Wilkerson C, Last RL, Pichersky E (2009) Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proc Natl Acad Sci USA **106:** 10865–10870

Schmidt A, Gershenzon J (2008) Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). Phytochemistry **69:** 49–57

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc **3:** 1101–1108

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ Jr (2005) Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol **6:** R33

Schwab W, Fischer TC, Giri A, Wüst M (2015) Potential applications of glucosyltransferases in terpene glucoside production: impacts on the use of aroma and fragrance. Appl Microbiol Biotechnol **99:** 165–174

Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, Chen H, Zeng J, Zhou Q, Wang S, Gu W, et al (2014) Biosynthesis, regulation, and domestication of bitterness in cucumber. Science **346:** 1084–1088

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13:** 2498–2504

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:** 2725–2729

Tholl D, Boland W, Hansel A, Loreto F, Röse USR, Schnitzler JP (2006) Practical approaches to plant volatile analysis. Plant J **45:** 540–560

Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E (2005) Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant J **42:** 757–771

Tholl D, Kish CM, Orlova I, Sherman D, Gershenzon J, Pichersky E, Dudareva N (2004) Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. Plant Cell **16:** 977–992

Tieman D, Bliss P, McIntyre LM, Blandon-Ubeda A, Bies D, Odabasi AZ, Rodríguez GR, van der Knaap E, Taylor MG, Goulet C, et al (2012) The chemical interactions underlying tomato flavor preferences. Curr Biol **22:** 1035–1039

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc **7:** 562–578

Wan H, Zhao Z, Qian C, Sui Y, Malik AA, Chen J (2010) Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber. Anal Biochem **399:** 257–261

Wang C, Chen Q, Fan D, Li J, Wang G, Zhang P (2016) Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in Brassicaceae plants. Mol Plant **9:** 195–204

Wang G (2014) Recent progress in secondary metabolism of plant glandular trichomes. Plant Biotechnol **31:** 353–361

Wang G, Dixon RA (2009) Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci USA **106:** 9914–9919

Wang G, Tian L, Aziz N, Broun P, Dai X, He J, King A, Zhao PX, Dixon RA (2008) Terpene biosynthesis in glandular trichomes of hop. Plant Physiol **148:** 1254–1266

Wang S, Ku SS, Ye X, He C, Kwon SY, Choi PS (2015) Current status of genetic transformation technology developed in cucumber (*Cucumis sativus* L.). J Integr Agric **14:** 469–482

Weitzel C, Simonsen HT (2015) Cytochrome P450-enzymes involved in the biosynthesis of mono- and sesquiterpenes. Phytochem Rev **14:** 7–24

Weng JK (2014) The evolutionary paths towards complexity: a metabolic perspective. New Phytol **201:** 1141–1149

Widhalm JR, Jaini R, Morgan JA, Dudareva N (2015) Rethinking how volatiles are released from plant cells. Trends Plant Sci **20:** 545–550

Xu H, Zhang F, Liu B, Huhman DV, Sumner LW, Dixon RA, Wang G (2013) Characterization of the formation of branched short-chain fatty acid:CoAs for bitter acid biosynthesis in hop glandular trichomes. Mol Plant **6:** 1301–1317

Yoo SD, Cho YH, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc **2:** 1565–1572

Zhang Z, Mao L, Chen H, Bu F, Li G, Sun J, Li S, Sun H, Jiao C, Blakely R, et al (2015) Genome-wide mapping of structural variations reveals a copy number variant that determines reproductive morphology in cucumber. Plant Cell **27:** 1595–1604