BMC Plant Biology

RESEARCH

Open Access

Loss-of-function variants of *CYP706A3* in two natural accessions of *Arabidopsis thaliana* increase floral sesquiterpene emission

Moonyoung Kang[1], Yuri Choi[1], Hyeonjin Kim[1], Min-Soo Choi[1], Seula Lee[2], Youbong Hyun[2*] and Sang-Gyu Kim[1,3*]

Abstract

Background The major floral scent compounds of *Arabidopsis thaliana* flowers are terpenes. Although *A. thaliana* is generally considered to be a self-pollinating plant, there are natural variation in terpene volatile emission from flowers. However, the genetic mechanisms underlying the natural variation in *Arabidopsis* floral scents remain limited.

Results Here, we screened 116 natural accessions of *A. thaliana* and observed a substantial variability in the levels of terpene emission across these accessions. A genome-wide association study (GWAS) uncovered a genomic region associated with the observed variability in myrcene, one of monoterpene compounds. We then performed high-throughput genetic mapping using two representative accessions: Col-0 and Fr-2, which emit low and large amounts of floral terpenes, respectively. Next-generation mapping and RNA sequencing analyses revealed that the natural premature stop codon of *CYP706A3* of Fr-2, located at the 98th codon, confers high emission of sesquiterpene from flowers. We also found an independent mutation of *CYP706A3* of Np-0 in different position, leading to increased sesquiterpene emission. Interestingly, the expression levels of defense-related genes in Fr-2 were lower than those in Col-0 flowers, which suggests that terpene volatiles are potentially linked to floral defense.

Conclusions The natural variation in *Arabidopsis* floral scent emission was partially explained by one natural allele of *CYP706A3*. Since some natural accessions harboring a functional allele of *CYP706A3* still emit the large amount of floral sesquiterpene, it is possible that rare variants located on other loci increase scent emission.

Keywords *AT5G44620* (*CYP706A3*), *AT5G44630* (*TPS11*), Cytochrome P450, Floral scent, Natural variation, Sesquiterpene, Terpene synthase

Background

Terpenes are the most diverse floral scent compounds in seed plants. These volatiles play essential roles in attracting pollinators, repelling florivores, and affecting microbe composition [1–3]. *Arabidopsis thaliana* emits terpenes as its major floral scents, and their biosynthetic enzymes have been elucidated since the early 2000s. For instance, terpene synthase 14 (TPS14) and TPS24 are the cyclases for producing monoterpenes including pinene or linalool [4]. TPS21 is responsible for synthesizing a subset of sesquiterpenes categorized as Group A: β-caryophyllene and humulene, while

*Correspondence:
Youbong Hyun
youbong.hyun@snu.ac.kr
Sang-Gyu Kim
sgkim1@kaist.ac.kr
[1] Department of Biological Sciences, Korea Advanced Institute for Science and Technology, Daejeon 34141, Republic of Korea
[2] School of Biological Sciences, Seoul National University, Seoul 08826, Republic of Korea
[3] Graduate School of Engineering Biology, Korea Advanced Institute for Science and Technology, Daejeon 34141, Republic of Korea

TPS11 synthesizes sesquiterpenes categorized as Group B: barbatene, *cis*-thujopsene, and β-chamigrene [4, 5]. The terpene can be oxidized to soluble terpene by cytochrome P450 (CYP) enzymes [6, 7]. *TPS* and accounting *CYP* genes are often clustered in close proximity [8, 9]. In *A. thaliana*, *TPS11* and *CYP706A3* are clustered within five kilobases. CYP706A3 was found to suppress the emission of sesquiterpene and monoterpene by solubilizing terpenes into oxidized form [7].

Although *A. thaliana* is commonly assumed as a self-pollinating plant, their flowers are visited by several insects in the natural populations [10]. Terpene oxides deter insect larvae from feeding on *A. thaliana* flowers [6]. Furthermore, the amount of terpene and its oxides is related to floral microbial populations. Loss of *CYP706A3* is associated with shift in bacterial communities [7], and loss of *TPS21* in *A. thaliana* eliminate the major volatile (*E*)-β-caryophyllene from flowers, leading to greater microbial pathogen infection on stigma [3]. Likewise, in *Solanaceae* flowers, inter-organ terpenoid transport, which is called natural fumigation, was also proposed to protect reproductive organs against microorganisms [11].

To understand intraspecific phenotypic differences and their fitness value in nature, genetic variation in *A. thaliana* accessions has been investigated since 1990s [12–14]. Variation in floral scent profiles also have been studied among accessions; for example, flower volatiles of 37 accessions were surveyed [5]. Herbivory-induced volatile emissions varied among 27 accessions [15], which uncovered non-functional alleles of terpene synthases. Recently, the 1001 Genomes Consortium deposited the genomes of 1,135 *A. thaliana* accessions, providing a valuable resource for various genome-wide analyses [16–18].

In this study, to identify genetic variants determining intraspecific phenotypic variation of floral volatile amounts, we screened flower scent profiles of over 100 natural accessions of *A. thaliana*. The analysis of genome-wide association study (GWAS) identified the loci associated with myrcene and additional non-significant loci related to Group B sesquiterpene emission. To identify loci underlying the amount of Group B sesquiterpenes, we constructed F_2 mapping populations between representative accessions Col-0 and Fr-2, which emit none or all of Group B sesquiterpene, respectively. Next-generation mapping and RNA sequencing analyses show that the natural loss-of-function allele of *CYP706A3* induces the high levels of Group B sesquiterpene emission from Fr-2 flowers. We also identified that a similar non-functional allele of *CYP706A3* occurred in another accession that emits large amount of sesquiterpenes from flowers, suggesting that this is one of the natural alleles that increase floral terpene emission in *Arabidopsis*.

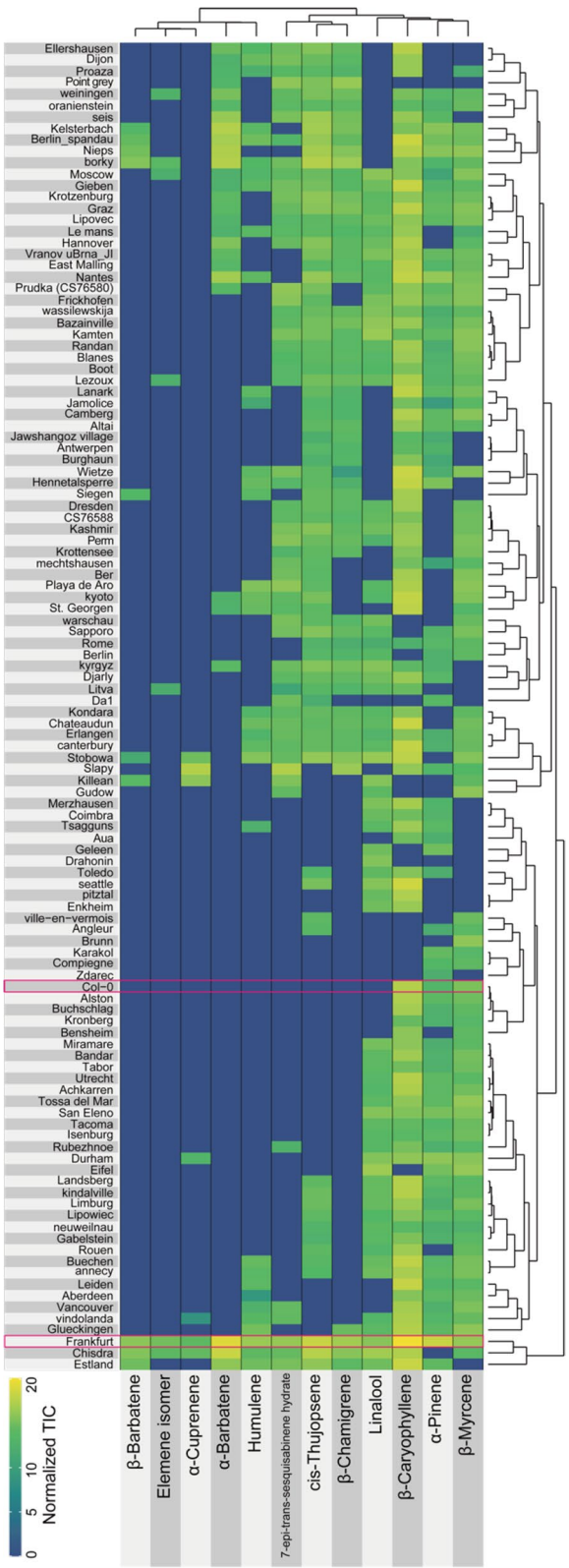
Results

Fr-2 has the most scented flowers among *A. thaliana* accessions

To assess intraspecific phenotypic variation of *A. thaliana* floral volatiles, we analyzed flower headspace volatile profiles over 100 accessions. Floral scents were collected from four to five flowers using silicone tubing located above the flowers. The result showed that (*E*)-β-caryophyllene, known as a major flower volatile of *A. thaliana* flowers, was detected from most accessions; approximately 90% of the accessions emit this sesquiterpene (Fig. 1). Pinene and myrcene were also detected from more than 70% of the accessions, followed by linalool and cis-thujopsene, which were detected from approximately 60% of the accessions. We found that some accessions emit all reported types of terpenes, while others emit only a subset or produce them at low intensity; for instance, Fr-2 flowers produced all reported sesquiterpenes at detectable levels. Although Col-0 has been chosen as a reference accession in several studies regarding floral terpenes, its flower was one of the least scented. Majority of Group B sesquiterpenes such as barbatene and *cis*-thujopsene were detected in samples collected from several accessions, but these terpenes were not well-detected in Col-0 flowers in our analysis. In addition, sesquiterpenes were not detected in some accessions, such as Zdarec, Brunn, and Drahonin. This suggests that these accessions either do not emit sesquiterpene, or produce limited amounts of sesquiterpenes that cannot be easily detected using the same procedure.

The Group B sesquiterpene phenotype observed in Fr-2 is inherited in a recessive manner

To identify the locus associated with terpene emission, we firstly performed genome-wide association analyses via easyGWAS [18]. We found that a single nucleotide polymorphism (SNP) upstream of *TPS21* is weakly associated with (*E*)-β-caryophyllene emission (Fig. S1), though the association is not statistically significant. This suggests that few accessions may exhibit changes in the activity of TPS21, its transcript levels, or the expression of nearby genes. We also found two genomic regions weakly but not significantly associated with pinene and five genomic regions significantly related to myrcene emission (Figs. S2, S3). However, those regions did not include any promising candidates within 100 kilobase intervals. In addition, we could not find any loci significantly associated with the emission of linalool and *cis*-thujopsene (Figs. S4, S5). Therefore, we designed alternative strategy to identify



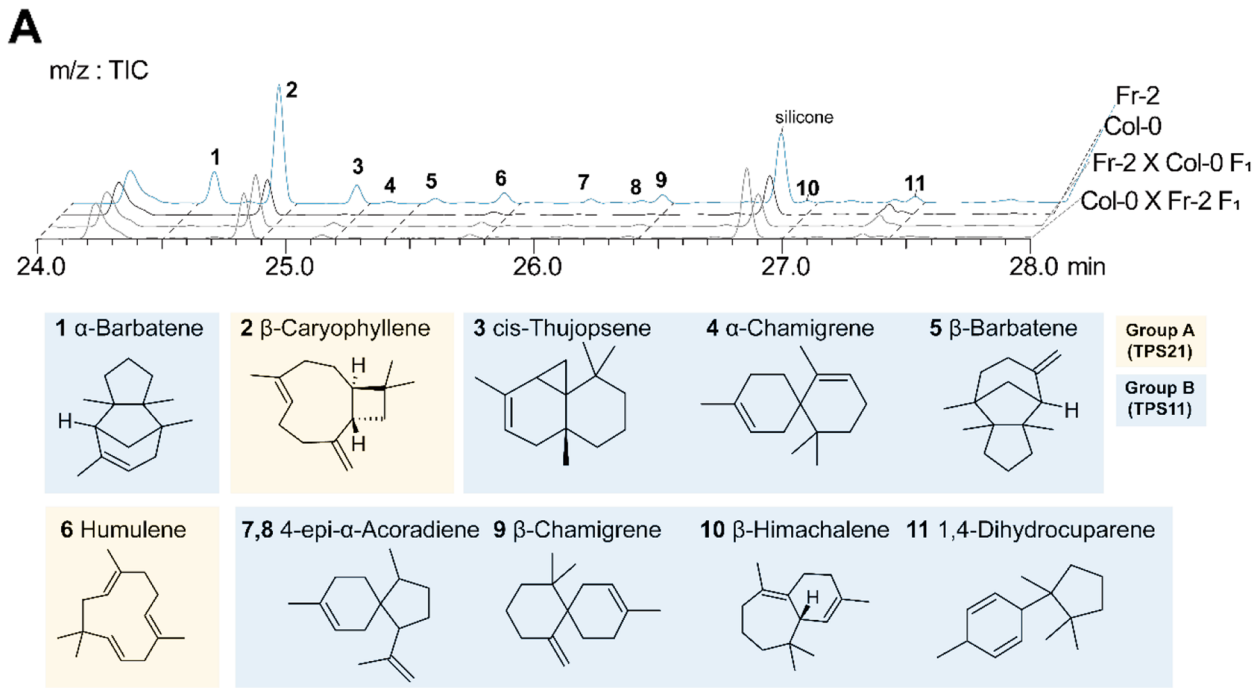
◀ **Fig. 1** Screening flower scent profiles of 116 *A. thaliana* accessions. The heatmap shows flower scent profiles from 116 *A. thaliana* accessions, involving all groups of terpenes. Color gradients reflects total ion counts (TIC), which were normalized by the number of flowers used for the experiment. Red boxes denote two accessions, Frankfurt (Fr-2) and Columbia (Col-0), which respectively represent a terpene-rich and a terpene-scarce group in this study

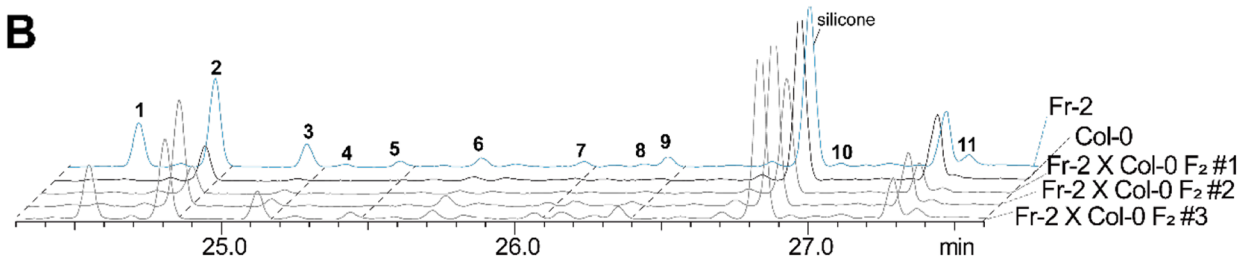
the causal genomic regions that explain the variation of Group B sesquiterpenes including *cis*-thujopsene. We chose Fr-2 as a representative accession among Group B sesquiterpene-rich accessions and Col-0 as a representative accession emitting less sesquiterpenes. We generated F_2 mapping population for next generation mapping (NGM) studies by crossing those two accessions [19].

Flowers of all F_1 plants emitted volatiles similar to Col-0 flowers, indicating that Fr-2-like scent profile seemed to be recessive (Figs. 2A, C). We propagated F_1 plants to the F_2 generation, and massively screened over 200 plants for measuring their floral volatiles. Fr-2- and Col-0-like phenotypes were segregated. As presented in Fig. 2B, for example, some F_2 plants showed Col-0-like scent profiles (F_2 #1), but others were similar to Fr-2 (F_2 #3). The genetic loci affecting amount of Group A and B sesquiterpenes appeared to be unlinked, because some F_2 plants emitted large amount of Group A sesquiterpene and, at the same time, less amount of Group B sesquiterpene (F_2 #2). Regarding the Group B sesquiterpene emission, the ratio of Col-0-like to Fr-2-like phenotype was close to 3:1 in the F_2 generation, suggesting that this phonotypic variation is determined by a single genetic locus or tightly linked loci (Fig. 2C).

The difference of Group B sesquiterpene emission between Fr-2 and Col-0 flowers is associated with the *TPS11-CYP706A3* gene cluster

As Group B sesquiterpene emission was linked with one locus, we adopted NGM pipeline to map candidate causative region (Fig. 3A). Genomic DNA was extracted from F_2 individuals which emitted large amount of Group B sesquiterpenes, and the pooled samples were subjected to Illumina sequencing. Paired-end reads were mapped to the TAIR10 release of the *Arabidopsis* Col-0 genome, then SNPs were called. The resultant vcf file was uploaded on the NGM website and analyzed, revealing that Fr-2-like SNPs were enriched in the region on chromosome 5, where the *TPS11* (*AT5G44630*)-*CYP706A3* (*AT5G44620*) gene cluster is located, with calculated discordant chastity over 0.96 (Table S3). However, we found no detrimental SNPs, which might cause the premature stop codon in that region based on the public vcf files.





C

Category	Generation	No. of plants		Expected ratio	χ2	*P* (two-tailed)
		Col-0 like	Fr-2 like			
Col-0 X Fr-2	F_1	16	0	-		
	F_2	41	11	3:1	0.410	0.5218
Fr-2 X Col-0	F_1	16	0	-		
	F_2	133	33	3:1	2.321	0.1276

Fig. 2 Group B sesquiterpene phenotype of Fr-2 is recessive and segregated into 3:1 ratio in F_2 generation. **A** Flowers of F_1 plants emit similar scent profiles to Col-0 flowers, suggesting highly scented phenotype of Fr-2 flowers is recessive trait. Numbers denoted on chromatogram are floral sesquiterpenes. The names and chemical structures of sesquiterpenes matched to each number is presented. **B** The chromatograms show that some of F_2 plants exhibited Fr-2-like scent profiles while others emit like Col-0 flowers. **C** F_2 progenies resulting from crossing Col-0 and Fr-2 display a segregation ratio of 3:1 with respect to their floral bouquet profiles, thus following Mendelian laws of segregation. This was determined based on a two-sided chi-square test

We next hypothesized that the expression of candidate genes differs among *A. thaliana* accessions. To evaluate differentially expressed genes (DEG) in flowers, we analyzed floral transcriptome of the two accessions using RNA sequencing. 1,036 genes in Col-0 and 388 genes in Fr-2 were more expressed compared to the other (Fig. 3B, Table S4). Upregulated genes in Col-0 flowers were related to glucosinolate biosynthesis and plant-pathogen

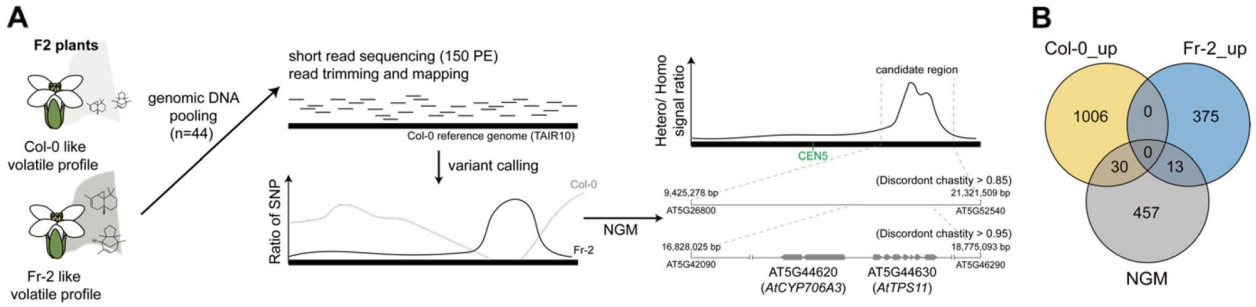


Fig. 3 Next generation mapping (NGM) pipeline designated the causal genomic region, which contains *TPS11-CYP706A3* gene cluster. **A** Schematics of NGM analysis conducted in this study. Genomic DNA was extracted from F_2 progenies showing Fr-2-like volatile profile and sequenced. The candidate region was located in chromosome 5, which contains around five hundred genes including the *TPS11-CYP706A3* gene cluster as illustrated. **B** To compare the overall transcriptome of two accessions, RNA sequencing was performed on Col-0 and Fr-2 flowers. A Venn diagram revealed that 1,036 genes were expressed more in Col-0 flowers, while 388 genes were expressed more in Fr-2 flowers. Among the genes located within the candidate region identified by the NGM pipeline, thirty genes and thirteen genes were expressed more in Col-0 flowers and Fr-2 flowers, respectively

interactions, while the genes upregulated in Fr-2 flowers were involved in photosynthesis and other metabolic pathways (Fig. S6A). Among the DEG, 43 genes were located in the candidate region identified by NGM pipeline (Fig. 3B, Table S4), and *CYP706A3* was one of them expressed less in Fr-2 flowers (Fig. S6B).

TPS11 in Fr-2 does not enhance sesquiterpene emission

As the *TPS11-CYP706A3* gene cluster was found in the causative region, we firstly checked whether the TPS11 of Fr-2 is more active than that of Col-0. Expression level of *TPS11* was similar between the two accessions (Fig. S6B), therefore we verified SNPs in *TPS11* by Sanger sequencing. As predicted in the 1001 genome vcf files, Fr-2 has 4 amino acid substitutions in TPS11, which could result in structural change (Fig. 4A).

To test whether Fr-2 TPS11 is more active in synthesizing Group B sesquiterpenes than Col-0 TPS11, we transiently expressed two forms of TPS11 in *Nicotiana benthamiana* and analyzed headspace volatiles from the infiltrated *N. benthamiana* leaf disks. Fr-2 TPS11 emitted slightly more sesquiterpene than Col-0 TPS11, but there was no significant difference (Fig. 4B). Additionally, we introduced each amino acid substitutions in Col-0 TPS11 and infiltrated into *N. benthamiana*, but neither of them improved the Col-0 enzyme efficacy (Figs. 4C, S7). Although those SNPs in *TPS11* seemed to be prevalent among other accessions based on vcf files (Figs. 4D, E), the existence of those SNPs was not correlated with sesquiterpene emission amount (Fig. 1). Therefore, we concluded that *TPS11* is not the causative gene that enhances sesquiterpene emission in Fr-2 flowers.

CYP706A3 in Fr-2 encodes a truncated enzyme, which increases sesquiterpene emission

We next examined whether *CYP706A3* is responsible for Group B sesquiterpene variations. RNA sequencing analysis revealed that Fr-2 flowers express less *CYP706A3*, which is a solubilizing oxidase for sesquiterpenes (Fig. S6B). There were numerous ambiguous bases in Fr-2 *CYP706A3* coding sequence in the public vcf file (Fig. 5A), thus we verified SNPs by Sanger sequencing (Fig. 5B). We found a 1 bp deletion in 234th base of Fr-2 *CYP706A3* gene among the total 1,560 bases of CDS, resulting in a premature stop codon. The low mRNA levels of *CYP706A3* in Fr-2 could be explained by the premature stop codon, which presumably activates nonsense-mediated mRNA decay pathway.

To confirm that the lack of CYP706A3 activity in Fr-2 flowers increases the emission of Group B sesquiterpenes, we expressed Col-0 version of *CYP706A3* in the Fr-2 background. We found that overexpressing $CYP706A3_{(Col\text{-}0)}$ significantly reduced the amount of floral sesquiterpene emission in four independent transgenic lines (Figs. 5C, S8). Therefore, expression of functional *CYP706A3* was sufficient to restore the Col-0 phenotype in the Fr-2 background, indicating that the natural loss of *CYP706A3* resulted in high sesquiterpene emission in Fr-2 flowers.

Another accession with high amount of floral sesquiterpenes lost *CYP706A3* function

As we found that the *CYP706A3* mutation increased floral scents in Fr-2 flowers, we next examined whether this phenomenon is prevalent among other accessions. We chose 8 accessions which emitted large amount of Group B sesquiterpene in the previous screening (one accession does not have public dataset), and firstly checked

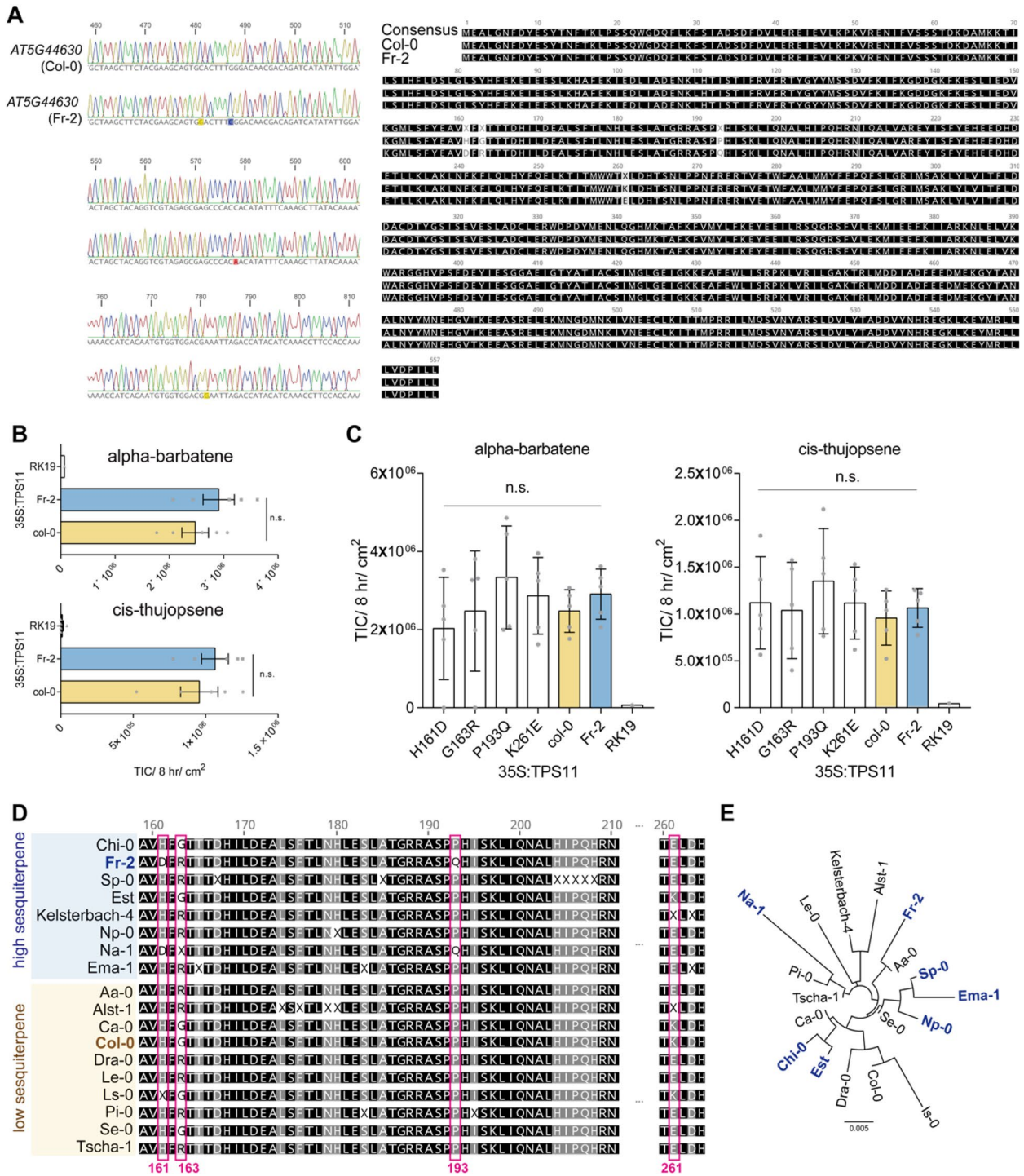


Fig. 4 The activity of TPS11 in Fr-2 is similar to that in Col-0. **A** Sanger sequencing of the coding sequence of *TPS11* confirms the SNPs previously characterized by the 1001 genome project. TPS11 of Fr-2 differs from that of Col-0 in four amino acids. **B** Leaf disks of *N. benthamiana* that were infiltrated with two forms of TPS11 emitted similar amounts of their major products, α-barbatene and *cis*-thujopsene. Two-tailed Student's *t*-test; n.s., not significant. **C** The efficacy of Col-0 TPS11 was not found to differ when amino acid residues were substituted with those of Fr-2. One-way ANOVA followed by Tukey's multiple comparisons; n.s., not significant. **D** Four amino acid variants were also found in other accessions, not correlated with the emission amounts of sesquiterpenes from their flowers. Red boxes denote the positions of amino acid residues, which are different from those of Col-0. **E** Phylogenetic trees constructed from protein sequences of TPS11 did not reveal any significant divergence between the two groups. The blue group represents accessions with high sesquiterpene emission levels

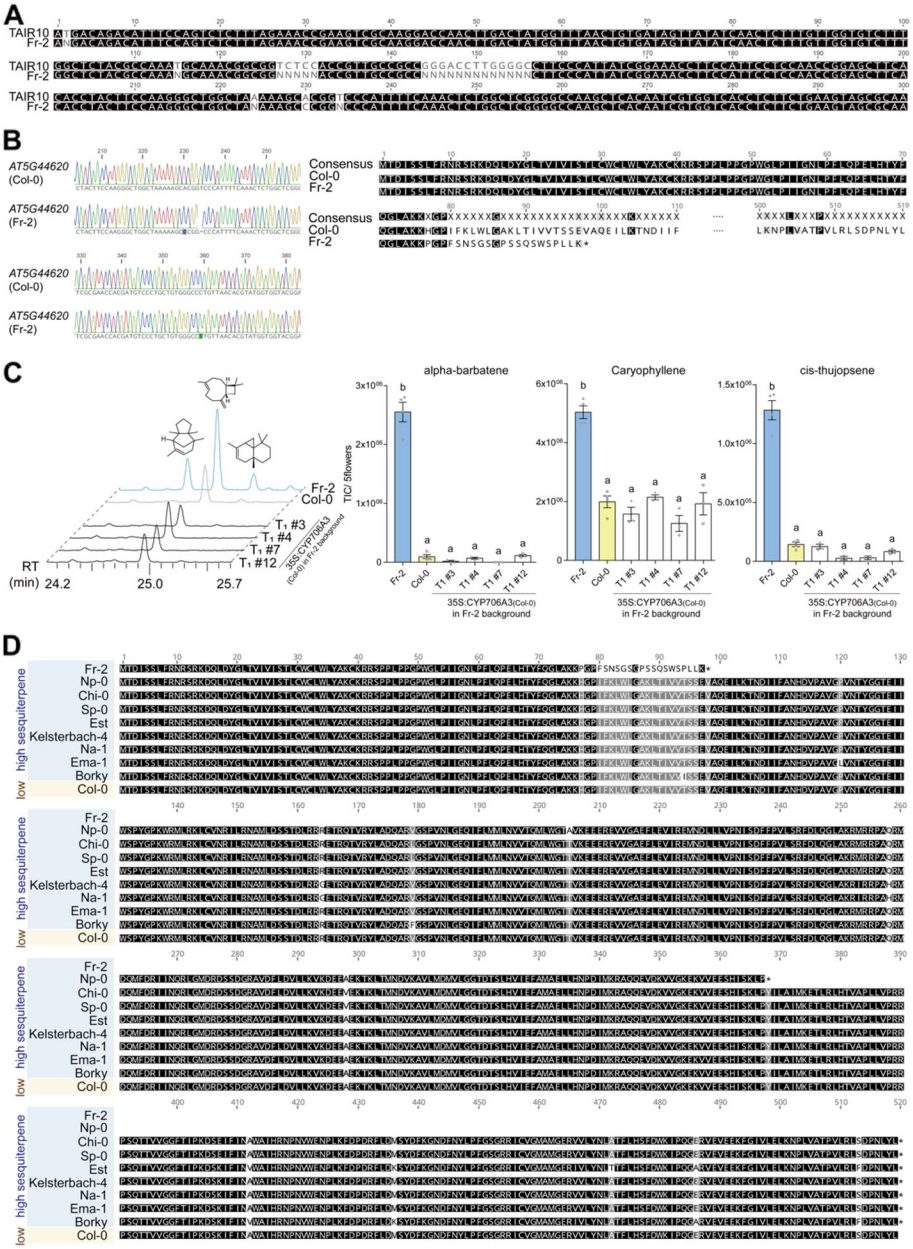


Fig. 5 Loss of *CYP706A3* increases floral sesquiterpene emission in two natural accessions. **A** *CYP706A3* coding sequence of Fr-2 harbored uncharacterized SNPs in the vcf file. **B** SNPs were verified by Sanger sequencing. A 1 bp deletion resulted in premature stop codon in *CYP706A3* of Fr-2. **C** Overexpression of *CYP706A3*(Col-0) in Fr-2 was sufficient to decrease sesquiterpene emission comparably to Col-0. One-way ANOVA followed by Tukey's multiple comparisons. **D** Accessions whose flowers emit sesquiterpene harbor SNPs in *CYP706A3* coding sequence. Nieps (Np-0) also has premature stop codon in 368th codon, verified by Sanger sequencing

their *CYP706A3* coding sequences in vcf files (Fig. S9). Among them, Np-0 was predicted to have a premature stop codon in this gene, and others were not. We further validated their sequences by Sanger sequencing (Fig. 5D), and found that Np-0 had a premature stop codon in *CYP706A3*.

Discussion

A. thaliana accessions that emit large amount of floral Group B sesquiterpene have mutations in *CYP706A3* gene

A. thaliana emits monoterpene and sesquiterpene from their flowers as well as vegetative tissues, which can serve as potential defensive molecules against pathogens or insects [3, 20]. Such metabolites involved in

ecological interactions are under selection pressures in various natural conditions, thus profiles of them tend to be diverse within a species. To elucidate the genetic basis of intraspecific variation in terpene emission, several research groups have used tens of *A. thaliana* natural accessions for decades. Absence of sesquiterpene in some accessions was reported to be resulted from mutation of *TPS* genes [5]. Additionally, insect-induced terpene emissions were found to vary among accessions. Investigation on those phenotypic variation helped characterizing loss-of-function alleles of two terpene synthases in some accessions, which resulted in lack of induced emission of terpene [15].

In this study, we analyzed flower volatiles over one hundred natural accessions of *A. thaliana*. We found that several accessions emit large amount of flower sesquiterpene compared to Col-0. GWAS analysis identified one SNP upstream of *TPS21* that was weakly, but not significantly, associated with the emission of (*E*)-β-caryophyllene, a major sesquiterpene of Group A. This suggests that the variability of caryophyllene emission, though not for all accessions, may be partially explained by this locus. In addition, some genomic regions were associated with monoterpene emission, but no promising candidate genes were identified in their vicinity. In our GWAS analysis, we were unable to identify loci associated with Group B sesquiterpenes, likely due to the incompleteness of the dataset. The very low levels of Group B sesquiterpenes in some accessions may not be detected due to the sensitivity constraints of mass spectrometry. Another possibility is that multiple rare loci may affect the scent trait, requiring a larger number of accessions to identify significantly associated loci. Since we were unable to specify the loci responsible for sesquiterpene emission via GWAS with the number of accessions used in the study, we performed next generation mapping experiments and RNA sequencing analyses. These approaches allowed us to identify a loss-of-function mutation of *CYP706A3* in Fr-2. We also identified another accession harboring the mutation in *CYP706A3* gene. This allele differed from that of Fr-2, suggesting that the mutated alleles emerged independently.

Fr-2 complemented with the functional allele of *CYP706A3* emitted less sesquiterpenes, which was comparable to Col-0 flowers. In addition, the Group A sesquiterpene such as (*E*)-β-caryophyllene also decreased in the *CYP706A3*-overexpressing Fr-2, suggesting that the loss of *CYP706A3* might increase other major floral sesquiterpenes (Fig. 5C). This result also supports the previous findings that CYP706A3 is the promiscuous oxidase [7]. However, (*E*)-β-caryophyllene emission of F_2 generation flowers were distributed continuously between two accessions (Fig. S10). This result indicates that (*E*)-β-caryophyllene emission could be a quantitative trait that is under the control of multiple loci, which may involve *CYP706A3* and the promoter region of *TPS21*.

We found that the *TPS11-CYP706A3* gene cluster emerged in the family *Brassicaceae* [7] was decoupled in two *Arabidopsis* accessions: Fr-2 and Np-0, resulting in more scent emission. However, we also observed that other accessions, which emit large amount of sesquiterpene, still retain functional CYP706A3. It suggests the existence of other mechanism boosting floral sesquiterpene emission. We hypothesize that there could be other redundant oxidases homologous to CYP706A3 or supplementing mechanisms for sesquiterpene production, which still remains to be investigated.

The global transcriptome of Fr-2 flowers suggests a potential connection between defense mechanisms and oxygenated sesquiterpenes

The reason why some of *Arabidopsis* accessions lost the coupling of *TPS11-CYP706A3* gene cluster remains further explored. It can be possible that the loss-of-*CYP706A3* allele increases the visibility of flowers to thrips, which can deliver pollens to other plants. In addition, our results indirectly support the previous hypothesis that terpene oxides in *A. thaliana* flowers serve as protections against visiting insects [6, 7]. In accordance with the latter hypothesis, Col-0 flowers with functional CYP706A3 globally expressed defense-related genes more than did Fr-2 flowers. Genes associated with plant-pathogen interactions and glucosinolate biosynthesis were highly expressed in Col-0 flowers. In contrast, Fr-2 flowers were likely to express more genes involved in core metabolic pathways, suggesting that there are some trade-off between synthesizing soluble defense molecules and core metabolism in *A. thaliana* flowers [21, 22]. However, further studies are required to evaluate the fitness of two accessions in the context with the flower visitors.

Conclusions

In summary, we investigated floral scent profiles of 116 *A. thaliana* accessions to find intraspecific variations on amounts of emission. We first performed GWAS to demonstrate genetic architecture of the scent variation; few genomic regions were found to be associated with (*E*)-β-caryophyllene and myrcene, but not for other sesquiterpenes. Through NGM analysis, we designated that *CYP706A3* is one of the loci that confers variability of floral scent amounts.

Methods

Plant growth

A. thaliana ecotype Columbia-0 (Col-0) and Frankfurt-2 (Fr-2) seeds were sown on half-strength Murashige and Skoog (MS) medium after sterilized with 75% ethanol supplemented with 0.05% Triton X-100. Seeds were stratified for 3 days in 4 ℃ dark, then grown in growth chamber for one week. Seedlings were transplanted into a soil mixture in 6 cm diameter pots, and cultivated in growth room under long day conditions (16-h light, 8-h dark cycle) at 25 ℃. *Nicotiana benthamiana* seeds were sown on a soil mixture and grown in the same condition for four weeks before infiltrated.

Flower headspace volatile analysis

We followed silicone tubing (ST) preparation and TD-GC–MS method as described in previous studies [23, 24]. Briefly, we prepared 1 cm long ST pieces (Cat. 9555.1, Carl Roth, Karlsruhe, Germany) which were tightly sealed in clean vials after baked. An inflorescence with four to five open flowers was excised from each *A. thaliana* ecotype, and was placed in a 2 mL clear vial containing 200 µL of distilled water. To collect the flower headspace volatiles, a 1 cm long ST piece was situated over the inflorescence for nine hours (from AM 11 to PM 8). STs were harvested after absorption, and stored in clean vials at −20 ℃. One day before TD-GC–MS analysis, STs were kept at room temperature for equilibration. STs were then placed into thermal desorption (TD) sampler tubes (Supelco) and volatiles were desorbed using a TD-30 thermal desorption unit (Shimadzu). Samples were then sent to a gas chromatography–mass spectrometry Nexis GC-2030–QP2020 NX (Shimadzu) for analysis. Gas chromatography unit was equipped with a SH-Rtx-5MS column (30 m long, 0.25 mm inner diameter, 0.25 µm film thickness, Shimadzu). GC separation and MS detection followed the oven program (40 °C for 5 min, ramp to 185 °C with 5 °C/min, ramp to 280 °C with 30 °C/min, hold for 0.83 min) and MS settings (240 °C interface, 220 °C ion source, full scan from 33 to 400 m/z) described in the previous method [23]. We identified sesquiterpenes based on their retention times and electron ionization mass spectra, which were compared with NIST library. We also referred to previously published *Arabidopsis* flower sesquiterpenes [4, 5, 7]. To normalize the quantity of volatiles, total ion counts of each compound were divided by the number of flowers, or the surface area of leaf disks, and the duration of scent absorption.

Next Generation Mapping (NGM) analysis

To map causative loci of Fr-2 phenotype, we followed NGM procedure described previously [19]. F_2 generation between Col-0 and Fr-2 ecotypes were generated and screened for their flower volatile phenotypes. Genomic DNA was extracted and pooled, from 44 individual plants which emitted less sesquiterpene from their flowers among the F_2 generation. DNA library was constructed using TruSeq DNA PCR-Free Kit (Illumina, San Diego, CA, USA) and sequenced based on Illumina platform (150 PE). Sequencing reads were quality-trimmed via cutadapt [25], resulting in adapter-free reads which were verified by fastqc [26]. Reads were mapped to the *A. thaliana* reference genome (TAIR10) by implementing BWA-MEM algorithm (v 0.7.17), and the resultant SAM file was converted to BAM file which was then sorted and indexed by the SAMtools (v 0.1.19) [27, 28]. We used SAMtools mpileup command to generate raw variant calls with the -E –ugf options, followed by BCFtools view command with –bvgN options to return the vcf file. This vcf file was processed into.ngm file based on a beta script provided by the previous study (http://bar.utoronto.ca/NGM/scripts/BCF2NGM.pl) [19]. The final.ngm file was uploaded and analyzed on the NGM website (http://bar.utoronto.ca/NGM/cgi-bin/emap.cgi), which revealed several SNP annotations enriched in chromosome 5.

RNA sequencing analysis

To assess differentially expressed genes between Col-0 and Fr-2 flowers, five open flowers (including sepals) were pooled per each replicate. Total RNA was extracted using RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) from two biological replicates. Sequencing libraries were constructed based on TruSeq Stranded Total RNA Library Prep Plant Kit following manufacturer's instructions, thereafter sequenced on Illumina platform (100 PE). Sequencing reads were trimmed via fastp [29], subsequently verified by fastqc. Trimmed reads were mapped to the *A. thaliana* TAIR10 reference genome (ENSEMBL annotation) with STAR aligner (v 2.7.9) [30]. Aligned reads were quantified by RSEM (v 1.3.0), and the resultant count data containing gene expression values were used for downstream analysis [31]. Tables of expected counts were used as inputs for edgeR, which identified differentially expressed genes by glm method after filtration and normalization [32]. Genes with false discovery rates (FDR) < 0.05 were considered significant, which were used for gene ontology (GO) enrichment analysis [33]. Functional enrichment analysis was conducted using g:Profiler with

default options, with default threshold (0.05) and g:SCS method [34].

Generation of vector constructs

For ectopic expression of genes under *35S* promoter, coding sequences (CDS) of *AtTPS11* or *AtCYP706A3* from each ecotype were introduced into pH2GW7-derived binary vector. For site-directed mutagenesis of *AtTPS11*, partial fragments of *AtTPS11* CDS were amplified using primers that create intended substitution and the amplified fragments were ligated to plasmid backbone based on Gibson assembly technique. For the detailed primer information used in this study, see Table S1. To generate transgenic plants, *Agrobacterium tumefaciens* (GV3101) harboring desired constructs were introduced into both ecotypes (Col-0 and Fr-2) using the floral dipping method [35, 36] without significant modifications. Briefly, open flowers from both ecotypes were removed the day of dipping, and the aerial part of plants were dipped in *Agrobacterium* (GV3101) cell suspension which harbors the desired construct, in 5% (w/v) sucrose solution supplemented with 0.02% Silwet L-77. Plants were then covered with plastic bag for one day, and after removal of the bag, they were grown to get seeds.

Nicotiana benthamiana infiltration for evaluating *AtTPS11* efficacy

Transient expression of *AtTPS11* in *N. benthamiana* was conducted as described previously [37]. Briefly, *A. tumefaciens* (GV3101) harboring constructs of interest or P19 protein were incubated at 28 ℃, respectively. When an OD_{600} reaches approximately 1.0 ~ 2.0, the cultured bacterial cells were harvested by centrifugation and resuspended in infiltration solution (10 mM MES pH 5.6, 10 mM $MgCl_2$, and 150 µM acetosyringone) to achieve OD_{600} 0.3 for P19 (RK19) and OD_{600} 0.5 for *AtTPS11* in mixed suspension. After left on bench for 2 h at room temperature, infiltration solution was injected into the abaxial side of a leaf using a 1 mL syringe. Plants were returned to growth room. At 3 days post-infiltration (dpi), a 9 mm diameter leaf disk (punched near the infiltrated scar) was placed in an 8 mL glass vial, and headspace volatiles emitted from the disk were trapped by a 1 cm long ST piece. ST pieces were located over the leaf disk for eight hours (from 11 AM to 7 PM). After volatile collection, leaf disks were harvested to extract total RNA using RNeasy Plant Mini Kit (Qiagen) following manufacturer's instruction. cDNA was synthesized using RNA to cDNA EcoDry Premix (Takara, Tokyo, Japan). To compare the expression level of each construct, relative transcript level of transiently expressed *AtTPS11* was determined by quantitative RT-PCR (2^-ΔΔCt method), using *NbEF1a* as a reference gene.

Comparing SNPs in *At5g44620* and *At5g44630*

Vcf files were downloaded from 1001 Genomes Project website, or each pseudogenome was downloaded from the website and parsed into FASTA file to contain GOI, which represent reference calls combined with variants including indels (https://1001genomes.org/data/GMI-MPI/releases/v3.1/) [16]. Since the nucleotide sequences contained several bases with uncertainty (N), their protein sequences were predicted assuming that the Ns encode the corresponding number of bases. To verify the sequence, we performed Sanger sequencing targeting CDS of each gene. Primers spanning CDS were designed and used to amplify the regions. PCR was conducted using Phusion DNA polymerase (Thermofisher, Waltham, MA, USA) in order to generate amplicons used for Sanger sequencing.

GWAS

Among 12 measured terpenes, five compounds that were detected from more than 50 accessions were selected for GWAS: alpha-pinene, beta-myrcene, beta-caryophyllene, linalool, and (*Z*)-thujopsene. The GWAS was performed on easyGWAS website (https://easygwas.ethz.ch/) using 1001 Genome dataset and TAIR 10 annotation [16, 18]. The normalized emission values were log10 transformed to be used as input phenotype values (Supplementary Table S2). EMMAX algorithm was selected for association mapping [38]. Loci that passed a Bonferroni-corrected threshold ($p = 0.05$) were considered to be significantly associated with the trait.

Abbreviations

CYP	Cytochrome P450
DEG	Differentially expressed genes
Fr-2	Frankfurt-2
Group A sesquiterpenes	Sesquiterpenes generated by AtTPS21
Group B sesquiterpenes	Sesquiterpenes generated by AtTPS11
GWAS	Genome-wide association study
NGM	Next generation mapping
Np-0	Nieps-0
SNP	Single nucleotide polymorphism
TPS	Terpene synthase
Vcf	Variant call format

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06283-6.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.

Acknowledgements
We appreciate Prof. Giltsu Choi, members of Laboratory for Ecology at KAIST.

Clinical trial number
Not applicable

Authors' contributions
S.-G.K. and Y.H. conceived the project and supervised the project. M.K., Y.H, and S.-G.K. wrote the manuscript. M.K. performed experiments and bioinformatics analysis. M.C. performed GWAS. Y.C., H.K., and S.L. conducted experiments. All authors reviewed the manuscript.

Funding
This work is supported by grants from RDA (RS-2024–00322407, RS-2024–00400556), Korea Polar Research Institute (KOPRI, PE23160), and TJ Park Doctoral Science Fellowship.

Data availability
Availability of data and materials The sequencing datasets used during the current study are deposited as a BioProject accession (PRJNA1004086) and will be released after acceptance. The reviewer link for the BioProject is https://dataview.ncbi.nlm.nih.gov/object/PRJNA1004086?reviewer=kh7kvbhkdb3qbbep8v44dq348, where reviewers can download the data during reviewing process. All other data generated during this study are included in this published article and supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 31 October 2024 Accepted: 20 February 2025
Published online: 03 March 2025

References

1. Reinhard J, Srinivasan MV, Zhang S. Scent-triggered navigation in honeybees. Nature. 2004;427:411.
2. Kessler D, Bing J, Haverkamp A, Baldwin IT. The defensive function of a pollinator-attracting floral volatile. Funct Ecol. 2019;33:1223–32.
3. Huang M, Sanchez-Moreiras AM, Abel C, Sohrabi R, Lee S, Gershenzon J, et al. The major volatile organic compound emitted from Arabidopsis thaliana flowers, the sesquiterpene (E)-β-caryophyllene, is a defense against a bacterial pathogen. New Phytol. 2012;193:997–1008.
4. Chen F, Tholl D, D'Auria JC, Farooq A, Pichersky E, Gershenzon J. Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. Plant Cell. 2003;15:481–94.
5. Tholl D, Chen F, Petri J, Gershenzon J, Pichersky E. Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant J. 2005;42:757–71.
6. Boachon B, Junker RR, Miesch L, Bassard JE, Höfer R, Caillieaudeaux R, et al. CYP76C1 (Cytochrome P450)-mediated linalool metabolism and the formation of volatile and soluble linalool oxides in arabidopsis flowers: A strategy for defense against floral antagonists. Plant Cell. 2015;27:2972–90.
7. Boachon B, Burdloff Y, Ruan JX, Rojo R, Junker RR, Vincent B, et al. A promiscuous CYP706A3 reduces terpene volatile emission from arabidopsis flowers, affecting florivores and the floral microbiome. Plant Cell. 2019;31:2947–72.
8. Field B, Osbourn AE. Metabolic Diversification — Independent Assembly of Operon-like Gene Clusters in Different Plants. Science (80-). 2008;320 April:543–7.
9. Liu Z, Cheema J, Vigouroux M, Hill L, Reed J, Paajanen P, et al. Formation and diversification of a paradigm biosynthetic gene cluster in plants. Nat Commun. 2020;11:1–11.
10. Hoffmann MH, Bremer M, Schneider K, Burger F, Stolle E, Moritz G. Flower Visitors in a Natural Population of Arabidopsis thaliana. Plant Biol. 2003;5:491–4.
11. Boachon B, Lynch JH, Ray S, Yuan J, Caldo KMP, Junker RR, et al. Natural fumigation as a mechanism for volatile transport between flower organs. Nat Chem Biol. 2019;15:583–8.
12. Murphy A, Taiz L. Comparison of metallothionein gene expression and nonprotein thiols in ten arabidopsis ecotypes: Correlation with copper tolerance. Plant Physiol. 1995;109:945–54.
13. Eichenberg K, Hennig L, Martin A, Schäfer E. Variation in dynamics of phytochrome A in Arabidopsis ecotypes and mutants. Plant, Cell Environ. 2000;23:311–9.
14. King G, Nienhuis J, Hussey C. Genetic similarity among ecotypes of Arabidopsis thaliana estimated by analysis of restriction fragment length polymorphisms. Theor Appl Genet. 1993;86:1028–32.
15. Huang M, Abel C, Sohrabi R, Petri J, Haupt I, Cosimano J, et al. Variation of herbivore-induced volatile terpenes among arabidopsis ecotypes depends on allelic differences and subcellular targeting of two terpene synthases, TPS02 and TPS03. Plant Physiol. 2010;153:1293–310.
16. Alonso-Blanco C, Andrade J, Becker C, Bemm F, Bergelson J, Borgwardt KMM, et al. 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell. 2016;166:481–91.
17. Ferrero-Serrano Á, Assmann SM. Phenotypic and genome-wide association with the local environment of Arabidopsis. Nat Ecol Evol. 2019;3:274–85.
18. Grimm DG, Roqueiro D, Salomé PA, Kleeberger S, Greshake B, Zhu W, et al. easyGWAS: A cloud-based platform for comparing the results of genome-wide association studies. Plant Cell. 2017;29:5–19.
19. Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, et al. Next-generation mapping of Arabidopsis genes. Plant J. 2011;67:715–25.
20. Gershenzon J, Dudareva N. The function of terpene natural products in the natural world. Nat Chem Biol. 2007;3:408–14.
21. Bekaert M, Edger PP, Hudson CM, Pires JC, Conant GC. Metabolic and evolutionary costs of herbivory defense: Systems biology of glucosinolate synthesis. New Phytol. 2012;196:596–605.
22. Züst T, Agrawal AA. Trade-Offs Between Plant Growth and Defense Against Insect Herbivory: An Emerging Mechanistic Synthesis. Annu Rev Plant Biol. 2017;68:513–34.
23. Mario K, Veit D, Eilers EJ, Schuman MC. Application of Silicone Tubing for Robust, Simple, High-throughput, and Time-resolved Analysis of Plant Volatiles in Field Experiments. Bio-Protocol. 2015;5:e1391.
24. Kallenbach M, Oh Y, Eilers EJ, Veit D, Baldwin IT, Schuman MC. A robust, simple, high-throughput technique for time-resolved plant volatile analysis in field experiments. Plant J. 2014;78:1060–72.
25. Martin M. Cutadapt removes adapter sequences from highthroughput sequencing reads. EMBnet J. 2011;17:10–2.
26. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
29. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.
30. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
31. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323. https://doi.org/10.1186/1471-2105-12-323.
32. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26:139–40.
33. Lee GY, Ham S, Lee SV. Brief guide to RNA sequencing analysis for nonexperts in bioinformatics. Mol Cells. 2024;47(5): 100060.
34. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47:W191–8.

35. Clough SJ, Bent AF. Floral dip: a simplified method forAgrobacterium-mediated transformation ofArabidopsis thaliana. Plant J. 1998;16:735–43.
36. Zhang X, Henriques R, Lin SS, Niu QW, Chua NH. Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method. Nat Protoc. 2006;1:641–6.
37. Rolland V. Determining the Subcellular Localization of Fluorescently Tagged Proteins Using Protoplasts Extracted from Transiently Transformed Nicotiana benthamiana Leaves BT - Photosynthesis: Methods and Protocols. In: Covshoff S, editor. New York. New York, NY: Springer; 2018. p. 263–83.
38. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010;42:348–54.

Publisher's Note