



HAL
open science

Transcriptomic analysis of deceptively pollinated *Arum maculatum* (Araceae) reveals association between terpene synthase expression in floral trap chamber and species-specific pollinator attraction

Mark Szenteczki, Adrienne Godschalx, Jérémy Gauthier, Marc Gibernau, Sergio Rasmann, Nadir Alvarez

► To cite this version:

Mark Szenteczki, Adrienne Godschalx, Jérémy Gauthier, Marc Gibernau, Sergio Rasmann, et al.. Transcriptomic analysis of deceptively pollinated *Arum maculatum* (Araceae) reveals association between terpene synthase expression in floral trap chamber and species-specific pollinator attraction. *G3*, 2022, 12 (9), pp.jkac175. 10.1093/g3journal/jkac175 . hal-03767083

HAL Id: hal-03767083

<https://hal.science/hal-03767083>

Submitted on 1 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TITLE

Transcriptomic analysis of deceptively pollinated *Arum maculatum* (Araceae) reveals association between terpene synthase expression in floral trap chamber and species-specific pollinator attraction

AUTHOR LIST

Mark A. Szenteczki^{1*}, Adrienne L. Godschalx¹, Jérémy Gauthier², Marc Gibernau^{3 +}, Sergio Rasmann^{1 +}, Nadir Alvarez^{2 4 +}

¹ Université de Neuchâtel, Institut de Biologie, Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland

² Geneva Natural History Museum, Route de Malagnou 1, 1208 Genève, Switzerland

³ CNRS – University of Corsica, Laboratory Sciences for the Environment (SPE - UMR 6134), Natural Resources Project, Vignola – Route des Sanguinaires, 20000 Ajaccio, France

⁴ Department of Genetics and Evolution, University of Geneva, Quai Ernest-Ansermet 30, 1211 Geneva, Switzerland

* Corresponding author: M.A. Szenteczki mark.szenteczki@unine.ch Tel. + 41 32 718 23 16

+ Authors participated equally and are considered joint senior authors

1. ABSTRACT

Deceptive pollination often involves volatile organic compound (VOC) emissions that mislead insects into performing non-rewarding pollination. Among deceptively pollinated plants, *Arum maculatum* is particularly well-known for its potent dung-like VOC emissions and specialized floral chamber, which traps pollinators – mainly *Psychoda phalaenoides* and *P. grisescens* – overnight. However, little is known about the genes underlying the production of many *A. maculatum* VOCs, and their influence on variation in pollinator attraction rates. Therefore, we performed *de novo* transcriptome sequencing of *A. maculatum* appendix and male floret tissue collected during- and post-anthesis, from ten natural populations across Europe. These RNA-seq data were paired with GC-MS analyses of floral scent composition and pollinator data collected from the same inflorescences. Differential expression analyses revealed candidate transcripts in appendix tissue linked to malodorous VOCs including indole, p-cresol, and 2-heptanone. Additionally, we found that terpene synthase expression in male floret tissue during anthesis significantly covaried with sex- and species-specific attraction of *Psychoda phalaenoides* and *P. grisescens*. Taken together, our results provide the first insights into molecular mechanisms underlying pollinator attraction patterns in *A. maculatum*, and highlight floral chamber sesquiterpene (*e.g.* bicyclogermacrene) synthases as interesting candidate genes for further study.

KEYWORDS

Chemical ecology; GC-MS; molecular ecology; plant-pollinator interactions; RNA-seq; sesquiterpenes

2. INTRODUCTION

Angiosperms have evolved to produce a wide array of specialized metabolites to mediate their interactions with other organisms and their environment, including volatile organic compounds (VOCs) emitted from flowers, leaves, fruits, and roots. While VOCs represent a relatively small proportion of all plant metabolomic diversity (Knudsen and Gershenzon 2020), they play important functional roles in defence against predators, pathogens, and abiotic stresses (Pichersky and Gershenzon 2002; Holopainen 2004; Irwin *et al.* 2004) and as cues for pollinator attraction (Pellmyr and Thien 1986; Knudsen and Tollsten 1993; Whitehead and Peakall 2009; Junker and Parachnowitsch 2015). Pollination, particularly by insects, appears to be a major driver of angiosperm diversification (van der Niet and Johnson 2012; Schiestl and Johnson 2013; Hernández-Hernández and Wiens 2020). Recent research has hypothesized, and partially demonstrated, that floral trait diversity is the result of complex interactions between plant genomic diversity among populations, pollinator network composition, and environmental conditions – all of which vary through space and time (Thompson *et al.* 2013, 2017; Friberg *et al.* 2019). Large molecular and ecological datasets with wide spatial and temporal coverage are therefore still needed, in order to further our understanding of the evolution of key traits such as floral scent.

Pollinator-mediated selection is known to influence floral morphology (Bröderbauer *et al.* 2013; Gervasi and Schiestl 2017) and colour (Schemske and Bradshaw 1999; Newman *et al.* 2012; Trunschke *et al.* 2021), and recent studies have also identified strong evidence for variation in VOC emissions driven by pollinator preferences. Notably, comparative community ecology data and gas chromatography – mass spectrometry (GC-MS) have been combined to identify convergent VOC bouquets in unrelated plant species with similar pollinators (Fenster *et al.* 2004; Schiestl and Johnson 2013; Junker and Parachnowitsch 2015), and divergent VOC bouquets in related species with different pollinators (Dobson *et al.* 1997; Urru *et al.* 2010;

Byers *et al.* 2014; Friberg *et al.* 2019). Moreover, the discovery of biochemical pathways underlying floral VOCs has greatly accelerated by next-generation sequencing (NGS) technologies such as mRNA sequencing (Dhandapani *et al.* 2017; Wong *et al.* 2017b; Xiao *et al.* 2019). Here, we aimed to combine these techniques, in order to understand how pollinator-mediated selection influences transcriptomic variation underlying floral scent divergence in deceptively pollinated *Arum maculatum* L. (Araceae).

Arum maculatum is a common European woodland flower with a long history of study due to its deceptive pollination strategy, which involves olfactory deception through brood-site mimicry (Schmucker 1925; Knoll 1926; Vogel 1965; Lack and Diaz 1991). Specifically, *A. maculatum* emit VOCs (*e.g.* 2-heptanone, indole, and p-cresol) during anthesis which are also present in cow manure (Kite 1995) and decomposing organic material (Gfrerer *et al.* 2021): the breeding substrates of *Psychoda phalaenoides* L. and *Psychoda (Psycha) grisescens* (Tonnoir) (Espíndola *et al.* 2011; Szenteczki *et al.* 2021). These two dipteran (Psychodidae) moth flies are trapped at varying frequencies across the species distribution of *A. maculatum* (Espíndola *et al.* 2011; Szenteczki *et al.* 2021; Gfrerer *et al.* 2021). Geographic variation in *A. maculatum* floral scent has also been observed, first in England (Kite *et al.* 1998; Diaz and Kite 2002) and France (Chartier *et al.* 2011, 2013), and recently, across most of the species distribution range (Szenteczki *et al.* 2021; Gfrerer *et al.* 2021), and may be linked to variation in pollinator attraction (Szenteczki *et al.* 2021; Gfrerer *et al.* 2021).

Arum maculatum VOC emissions during anthesis are highly variable, and also include several sesquiterpenes and other aliphatic, aromatic, monoterpene, and nitrogen-containing minor compounds which may also contribute to pollinator attraction (Kite 1995; Szenteczki *et al.* 2021; Gfrerer *et al.* 2021). Intra- and interpopulation VOC variation mainly centres around variation in the proportion of indole to terpenes (Chartier *et al.* 2013; Szenteczki *et al.* 2021), which are respectively produced by aromatic amino acid (AAA; *i.e.* phenylalanine, tyrosine,

and tryptophan) metabolism and terpene synthases (TPS) via the mevalonate or methylerythritol 4-phosphate (MEP) pathways. Experiments with scented bait traps in England have shown that indole, 2-heptanone, and p-cresol are attractive to *Psychoda phalaenoides* (Kite *et al.* 1998). Large-scale field studies have further demonstrated that sesquiterpenes may also play a role in differential attraction of *P. phalaenoides* and *P. grisescens* (Szenteczki *et al.* 2021) and influence fruit set size (Gfrerer *et al.* 2021).

Another recognizable feature of *A. maculatum* is its inflorescence morphology. Like other monoecious Araceae, *A. maculatum* produce densely clustered, unisexual male and female florets arranged along a central spike (the spadix) which is surrounded by a leaf-like bract (the spathe). The spathe surrounds the fertile flowers, creating a basal trap chamber with only a narrow opening at the top of the chamber, which is surrounded by a ring of hair-like sterile flowers. The scent of the floral chamber during anthesis appears to be dominated by one sesquiterpene, bicyclogermacrene (Kite *et al.* 1998), which was previously misidentified as germacrene B (Kite 1995). The *A. maculatum* reproductive cycle takes place over two days (Gibernau 2004): on the evening of the first day, the appendix (sterile apex of the spadix) is thermogenic, heating up to 15°C above ambient air temperature. Pollinators are attracted by VOCs emitted by the appendix, and fall into the trap chamber, where they crawl over the male and female florets. No rewards are provided to pollinators within the chamber. A small amount of stigmatic secretions have been observed, but the low (9 - 12.5%) sucrose content suggests that this fluid is not nectar (Lack and Diaz 1991), and more likely is produced to collect pollen grains (Paiva *et al.* 2021). The following morning, after the female florets are no longer receptive, pollen is released onto the trapped pollinators. Finally, pollinators are able to escape later in the day, as the sterile hairs at the opening of the trap chamber begin to wither.

In closely related *Arum italicum*, sesquiterpene biosynthesis appears to begin in the male florets several days prior to anthesis, but the full range of characteristic floral VOCs such as p-cresol

and skatole are only detected on the day of anthesis, in appendix tissue (Leguet *et al.* 2014). To date, the biosynthetic pathways underlying the production of p-cresol and skatole have only been characterized in bacteria (Selmer and Andrei 2001; Liu *et al.* 2018), and it is unclear whether *A. maculatum* and other angiosperms use similar aromatic amino acid fermentation pathways to produce these VOCs. Furthermore, tissue-specific RNA-seq of *Arum concinatum* identified more diverse transcript expression in male floret tissue on the day of anthesis rather than in the appendix (Onda *et al.* 2015), even though *A. concinatum* also emits dung-like VOCs from its appendix (Urru *et al.* 2010). Consequently, many questions remain regarding the specific biosynthetic pathways underlying *A. maculatum* VOCs, the localization of their expression, and whether mRNA expression related to VOC production varies with geographic distance or pollinator community composition.

Here, we aim to address each of the aforementioned gaps in our knowledge. Our specific objectives were to: i) investigate variation in *A. maculatum* transcript expression between appendix and male floret tissue during anthesis, ii) use these results to identify differentially expressed transcripts putatively related to VOC biosynthesis, and iii) to characterize and compare differential expression in these transcripts across Europe, with a focus on populations with divergent pollinator attraction patterns. We predicted that during anthesis, aromatic amino acid (AAA) metabolism would be highly expressed in appendix tissue, while sesquiterpene synthases would be highly expressed in male floret tissue. We further expected to observe differential expression of transcripts associated with AAA metabolism and terpene synthases in inflorescences with dung-like (*i.e.* indole, 2-heptanone, and p-cresol) versus sesquiterpene-dominated floral bouquets, respectively. Finally, given that the *A. maculatum* population in Forêt du Gâvre, France appears to attract almost exclusively *P. grisescens* (Espíndola *et al.* 2011; Szenteczki *et al.* 2021), we predicted that putative locally adapted transcripts associated with particular VOCs may exist within this population. To test these predictions, we surveyed

VOC variation, pollinator attraction, and mRNA expression in both appendix and male floret tissue, across most of the *A. maculatum* species distribution range.

3. MATERIALS AND METHODS

3.1 – Sampling design

We sampled ten natural populations of *Arum maculatum* between April and May 2019 (Figure 1): three in France (Forêt du Gâvre, Conteville, and Chaumont), one in Switzerland (Neuchâtel), two in Italy (Montese and Rifreddo), one in Croatia (Visuč), two in Serbia (Gostilje and Sokobanja), and one in Bulgaria (Chiflik). Full details on sample data and population coordinates are given in Appendix S1, Table S1.

Before collecting tissue samples, we non-invasively sampled the dynamic headspace VOCs emitted by each *A. maculatum* inflorescence, following the methods detailed in (Szenteczki *et al.* 2021). Briefly, this involved placing polydimethylsiloxane (PDMS) coated Twister® stir bars (Gerstel) in glass tubes near the appendices of *A. maculatum* inflorescences on the evening of the day of anthesis (18:00 – 19:00), and pumping air over them for 30 minutes at an air flow rate of 200 mL min⁻¹. These samples were then kept on ice in individual sealed glass containers until GC-MS analyses, where VOCs were thermally desorbed and separated on a HP-5MS column. Major ions were recorded for each integrated peak using Agilent Chemstation, and compound identifications were derived from NIST 2.3 (library v.17) and published Kovats retention indices; all names used in our analyses should therefore be considered hypotheses. Appendix temperatures were recorded at the time of VOC sampling using FLIR® thermographic imaging, and ambient air temperature was recorded using a thermistor.

Immediately following VOC sampling (*i.e.* during anthesis), we used sterile scalpel blades to collect 1 mm thick slices of appendix and/or male floret tissue. As a control, male floret tissue samples were also collected from inflorescences the following morning, after pollinator attraction had ended. All samples were individually preserved in RNAlater® (Thermo Fischer

Scientific), and stored at -80°C until extraction. After collecting each during-anthesis male floret tissue sample, we closed and sealed the small window we cut in floral chamber, so that pollinator trapping was not affected. During the morning following anthesis, we collected all insects trapped within inflorescences, preserved them in 70% ethanol, and identified them to at least the suborder level. Psychodidae were further identified to species level (including sex) following taxonomic descriptions and illustrations (Ježek 1990).

3.2 – RNA-seq library preparation

All samples and tissue types were extracted using identical methods; approximately 100 mg of tissue was rinsed in sterile RNase-free water, flash frozen in liquid nitrogen, powdered in a QIAGEN TissueLyser II, and immediately extracted using a QIAGEN RNeasy Plant Mini Kit, following the manufacturer's protocol. Extraction quality and concentration were verified using an Agilent Fragment Analyzer. Three Illumina TruSeq® Stranded mRNA polyA libraries were prepared from these samples, following the manufacturer's recommended protocol. The first library contained the appendix RNA samples, and the second and third each contained a mix of during- and post-anthesis male floret tissue samples from all populations. Each library was then sequenced on its own lane (*i.e.* 3 lanes in total) using an Illumina HiSeq 4000 (150bp paired-end sequencing) at the Lausanne Genomic Technologies Facility (Switzerland).

3.3 – Preprocessing raw read data

We performed an initial quality filtering of all raw read files using fastp (Chen *et al.* 2018) to remove adapter sequences and polyA tails, trim reads with phred quality values below 30, and remove reads with >1 'N' bases. Then, we used rCorrector (Song and Florea 2015) to remove reads containing erroneous kmers (25-mers). Finally, we discarded unfixable read pairs following quality filtering with *FilterUncorrectablePEfastq.py* (<https://github.com/harvardinformatics/TranscriptomeAssemblyTools>).

3.4 – De novo transcriptome assembly, annotation, and expression quantification

We merged all of the quality filtered reads, and assembled a single *de novo* reference *A. maculatum* transcriptome (*i.e.* from all samples, tissue types, and populations) using Trinity v2.11.0 (Grabherr *et al.* 2011) with in-silico normalization, a *k*-mer size of 25, minimum contig size of 200bp, and strand-specific assembly enabled. Then, we tested two methods to filter out redundant transcripts with poor coding potential (*i.e.* minor isoforms resulting from incomplete assembly, sequencing errors, and/or heterozygosity introduced by our wide geographic sampling range): i) CD-HIT (Li and Godzik 2006; Fu *et al.* 2012) clustering at 95 percent identity with a word size of 5, and ii) the *tr2aacds* v4 pipeline from EvidentialGene (Gilbert 2019). We calculated assembly scores and assessed transcriptome completeness of both methods with BUSCO v5.0.0, using the Embryophyta_odb10 dataset (Seppey *et al.* 2019). Next, we used dammit v1.2 (Scott *et al.* 2019) to annotate our deduplicated assembly. This pipeline uses Transdecoder (<https://github.com/TransDecoder/TransDecoder>) to identify candidate coding regions, and searches the OrthoDB, Pfam-A, Rfam, and uniref90 protein databases for transcript annotations, with an E-value threshold of 1×10^{-5} . Unannotated transcripts were not removed from the final *A. maculatum* reference transcriptome. We also uploaded the EvidentialGene output to the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (Moriya *et al.* 2007) to generate KEGG Orthology (KO) annotations and pathway maps. Finally, we mapped the cleaned raw reads of each sample to our annotated reference transcriptome using Bowtie2 v2.3.5.1 (Langmead and Salzberg 2012), and quantified transcript expression using RSEM v1.3.2 (Li and Dewey 2011).

3.5 – Differential expression analyses

RSEM outputs were imported into R v.4.1.0 (R Core Team 2021) using *tximport* 1.20 (Soneson *et al.* 2015). Transcripts expressed below 0.75 counts per million (CPM) in at least seven samples (*i.e.* the smallest sample size for a group in our analyses) were then filtered out prior to subsequent analyses in DESeq2 1.3.2 (Love *et al.* 2014), using the *apeglm* log2 fold

change shrinkage estimation algorithm (Zhu *et al.* 2019). Transcripts with FDR-corrected p -values < 0.05 and \log_2 fold changes greater than 1.0 or less than -1.0 were considered as significantly differentially expressed among groups. We further identified significantly enriched gene sets in each tissue type using Gene Ontology (GO) term analysis with the R package *TopGO* (Alexa and Rahnenfuhrer 2010), and visualized the result using the R package *rrvgo* 1.4 (Sayols 2020), an implementation of the semantic similarity-based GO summary tool REViGO (Supek *et al.* 2011).

We performed three sets of differential expression analyses, each addressing one of the main aims of our study. Our first comparison was between appendix ($n = 7$) and male floret ($n = 16$) tissue collected during anthesis (*i.e.* at the same phenological time point). Second, we compared male floret tissue collected during anthesis ($n = 14$) against their paired control samples from the morning following anthesis ($n = 14$). Paired control samples were not collected for two inflorescences from Sokobanja, Serbia; these individuals were excluded from this analysis. Finally, we compared transcript expression in during-anthesis male floret tissue sampled in the Forêt du Gâvre population against all other populations, to identify transcripts putatively linked to this population's exclusive attraction of *P. griseus*. In order to account for some of the effects of isolation by distance, we further split this analysis along the two main neutral genetic clusters (Espíndola and Alvarez 2011): specifically, we compared during-anthesis male floret tissue from Forêt du Gâvre ($n=5$) against during-anthesis male floret tissue from i) France, Switzerland, and Italy ($n= 7$) and ii) Serbia ($n = 4$).

3.6 – Identifying transcripts underlying VOC biosynthesis

Building on the results of our differential expression analyses, we characterized and compared the expression of metabolic pathways underlying key *A. maculatum* VOCs known to be involved in the attraction of Psychodidae pollinators (Kite *et al.* 1998): indole, *p*-cresol, 2-heptanone, and sesquiterpenes. First, we reconstructed entire metabolic pathways using

automated annotations, GO terms, and KEGG pathway maps generated from i) our complete reference transcriptome and ii) sets of significantly differentially expressed genes. Since the genes responsible for the production of 2-heptanone and p-cresol in *A. maculatum* are unclear and unknown respectively, we also used homology searches to identify additional candidate genes. Here, we performed BLASTp searches of genes known involved in the biosynthesis of the aforementioned VOCs in other plants and bacteria, against a database of all of our Transdecoder predicted peptide sequences using, with an e-value cut-off of 1×10^{-5} . After all candidate transcripts were identified, we visualized variation in their average expression across all tissue types and populations using the R package *heatmap* (Kolde 2019)

3.7 – Coinertia analysis – terpene synthases correlated with VOCs and pollinator attraction

We extracted the coding sequences and DESeq2 normalized expression of all terpene synthases in our transcriptome (*i.e.* using uniref annotations, GO/KO terms, and PFAM domains), performed a multiple sequence alignment of these sequences using Clustal Omega (Sievers *et al.* 2011), created a Maximum Likelihood phylogeny using automatic substitution model selection and default parameters in IQ-TREE (Nguyen *et al.* 2015), and visualized the result using iTOL v5 (Letunic and Bork 2021). Then, we investigated whether terpene synthase expression was correlated with a) total terpene emissions during anthesis, or b) the composition (*i.e.* sex and/or species) of trapped pollinators within inflorescences. Specifically, we performed coinertia analysis using the R packages *vegan* (Oksanen *et al.* 2020), *RVAideMemoire* (Hervé 2021), and *ade4* (Thioulouse *et al.* 2018), comparing the TPS expression matrix produced above to corresponding matrices of i) proportional emissions of terpene VOCs ii) proportions of Psychodidae species trapped by the same *A. maculatum* inflorescences. Prior to this analysis, the gene-level TPS expression matrix was log-transformed and scaled, while the VOC and pollinator matrices were centred log ratio (CLR) transformed and scaled.

4. RESULTS

4.1 – Assembly and annotation quality

Between 27'933'352 and 66'825'898 (median: 40'354'964) paired-end reads from individual tissue samples passed all quality filtering steps (Appendix S1, Table S1). These were subsequently assembled into 273'346 Trinity genes, comprising 593'392 transcript isoforms. Trinity quality statistics (*e.g.* N50 of 1699 bp) are given in Appendix S1, Table S2. EvidentialGene filtered out a larger number of transcripts from this raw assembly (Appendix S1, Table S3) while maintaining a higher BUSCO scores than CD-HIT clustering (*e.g.* 95.6% complete versus 92.2% complete; full result in Appendix S1, Table S4). The EvidentialGene output was therefore annotated and used for all subsequent analyses. A total of 18'411 unique annotations were generated. 17'435 annotations mapped to Eukaryota (94.7%); 12'676 of these annotations further mapped to Mesangiospermae (68.9%); a relatively small number of annotations mapped to bacterial genes (692 results; 3.8%).

A total of 49'779 transcripts remained in our expression matrix following pre-filtering in DESeq2 (*i.e.* >0.75CPM in at least seven samples). Principal components analysis of filtered transcript expression (Appendix S1, Figure S1) revealed highly divergent transcript expression among male floret and appendix tissue during anthesis along the first PCA axis. Along the second PCA axis, samples were further divided between two main geographic regions – namely, northern populations (*i.e.* France, Switzerland, and northern Italy) and Balkan populations (*i.e.* Croatia, Serbia, and Bulgaria). This regional split is consistent with the two main genetic clusters identified using neutral (AFLP) markers (Espíndola and Alvarez 2011).

4.2 – Differential transcript expression between appendix and male floret tissue

Differential expression analysis comparing appendix and male floret tissue (both collected during anthesis) revealed 8683 transcripts with significantly greater expression in appendix tissue, and 6581 transcripts with significantly greater expression in male floret tissue (FDR

corrected p-values <0.05 ; log₂fold change ± 1), after controlling for population effects. Furthermore, we found that VOC biosynthetic activity is significantly elevated in *A. maculatum* appendix tissue during anthesis. Many major biosynthetic pathways (e.g. tryptophan/indole synthesis, terpene synthesis, and phenylpropanoid synthesis) were significantly enriched in the appendix tissue during anthesis, whereas transcripts related to pollen production were significantly enriched in male floret tissue (Figure 2).

4.3 – Differential transcript expression in male florets during- versus post-anthesis

When comparing during- and post-anthesis male floret tissue samples in a differential expression analysis that incorporated the paired nature of these samples, we identified 3847 transcripts with significantly greater expression during anthesis, and 2920 transcripts with significantly greater expression post-anthesis (FDR corrected p-values <0.05 ; log₂fold change ± 1). While tryptophan (*i.e.* indole) biosynthesis was elevated in male florets during anthesis, we did not identify increased expression of other putative VOC biosynthetic pathways (Figure 3). The full list of significantly enriched GO terms from both our first and second set of differential expression analyses are available in Appendix S2.

4.4 – Differential transcript expression associated with exclusive attraction of *P. griseus*

Our third and final set of differential expression analyses, comparing male floret tissue during anthesis among populations, revealed that i) 84 and 9 transcripts were respectively differentially expressed in Forêt du Gâvre versus all other French, Swiss, and northern Italian populations (FDR corrected p-value <0.05 ; log₂fold change ± 1) and ii) 327 and 175 transcripts were respectively differentially expressed in Forêt du Gâvre versus Serbian populations (FDR corrected p-value <0.05 ; log₂fold change ± 1). However, no transcripts putatively linked to VOC biosynthesis were identified among these transcripts.

4.5 – Candidate genes underlying *A. maculatum* floral VOCs

Consistent with recent large-scale surveys of *A. maculatum* floral scent (Szenteczki *et al.* 2021; Gfrerer *et al.* 2021) we observed considerable within-population variation in VOC bouquet composition among our samples (Appendix S1, Figure S2 and Table S5). We further identified candidate genes putatively linked to the biosynthesis of several dung-mimicking *A. maculatum* VOCs. First, we identified transcripts in both appendix and male floret tissue homologous with tryptophan synthase alpha subunit (TSA), which catalyses the conversion of indole-3-glycerolphosphate to indole.

Second, we identified homologs of two proteins which catalyse the degradation of tyrosine to 4-hydroxyphenyllactate: aromatic aminotransferase (ISS1) and hydroxyphenylpyruvate reductase (HPPR). Pearson correlation tests further identified several transcripts correlated with HPPR expression in appendix tissue during anthesis (Appendix S1, Table S6); notably, this included a putative dehydratase/shikimate dehydrogenase (Pearson $r = 0.988$, $p < 0.001$). This enzyme may further catalyse the conversion of 4-hydroxyphenyllactate to p-coumaric acid. However, we did not identify any homologs of bacterial proteins linked to the production of 4-hydroxyphenylacetate and/or p-cresol (Saito *et al.* 2018), such as hydroxyphenylacetate decarboxylase (4-Hpd).

Third, while phenylalanine-derived precursor molecules for several common floral benzenoid/phenylpropanoid (FBP) compounds – particularly p-coumaric acid – appear to be produced by *A. maculatum*, transcripts responsible for the production of benzenoid and phenylpropanoid VOCs generally appeared to be absent or weakly expressed in *A. maculatum*. Finally, we identified putative homologs of methylketone synthases (MKS1 and MKS2), which are known to be involved in 2-heptanone biosynthesis (Ben-Israel *et al.* 2009; Khuat *et al.* 2019). Heatmaps of the expression patterns for all of the aforementioned candidate transcripts are shown in Figure 4.

We identified *A. maculatum* transcripts putatively encoding all proteins in both the mevalonate and methyl-D-erythritol 4-phosphate (MEP) pathways. During anthesis, the MEP pathway appears to be more highly expressed than the mevalonate pathway, particularly in appendix tissue (Appendix S1, Figure S3). The first enzyme of the MEP pathway, 1-deoxy-D-xylulose-5-phosphate synthase (DXS), which was highly expressed in appendix tissue, also featured the greatest isoform diversity among all terpene backbone synthesis genes. Furthermore, we identified 108 putative TPS transcripts expressed in appendix and male floret tissue during anthesis. A phylogeny of all putative *A. maculatum* terpene synthases is given in Appendix S1, Figure S4, and expression patterns of TPS transcripts which passed our initial CPM filtering threshold are visualized in Figure 5.

Some of the putative terpene synthases we identified are homologs of proteins known to produce common *A. maculatum* VOCs (e.g. humulenes and germacrenes); others, such as 2-methylisoborneol (2-MIB) synthase, catalyse the production of terpenes not previously described in *A. maculatum*. However, in the case of the latter, the transcript we identified likely encodes a novel TPS, given that we identified an uncharacterized *Colocasia esculenta* (Araceae) gene which shares 92.5% identity with our transcript, as opposed to approximately 22.8% identity with the bacterial 2-MIB synthase (Appendix S1, Figure S5). Transcripts annotated as trimethyltridecatetraene synthase (Cyt P450 92C6) may represent another such novel TPS. This gene was the only putative VOC synthase we identified that was significantly overexpressed (FDR corrected p-values <0.05; log2fold change >1) in appendix tissue samples emitting high quantities of an unnamed sesquiterpene with a non-polar Kovats Retention Index of 1681, which is known to be a strong predictor of *P. grisescens* attraction (Szenteczki *et al.* 2021).

4.6 – Male floret terpene synthase transcript expression is correlated with pollinator attraction

Coinertia analyses did not identify covariation between putative TPS expression in male floret tissue, and proportional emissions of terpenes in VOC bouquets ($p = 0.964$, Monte-Carlo test, 999 replicates). However, we did identify significant concordance between putative TPS expression in male floret tissue and Psychodidae pollinator composition trapped within inflorescences ($p = 0.047$, Monte-Carlo test, 999 replicates); the first coinertia axis split TPS expression based on species (i.e. *P. grisescens* vs *P. phalaenoides*), while the second coinertia axis split TPS expression based on sex (Appendix S1, Figure S6). Consequently, while our coinertia analyses could not disentangle covariation in TPS expression and specific terpene VOC emissions, it appears that TPS expression in male florets is correlated with sex- and/or species-specific attraction of Psychodidae.

5. DISCUSSION

High-throughput transcriptome sequencing is rapidly advancing our understanding of the biosynthesis of floral volatile compounds. In this study, we were able to overcome some of the challenges associated with studying the genes underlying VOC biosynthesis in non-model systems (Wong *et al.* 2017b), by surveying transcript expression variation in multiple VOC-emitting floral tissues, across the Europe-wide distribution range of *A. maculatum* (Figure 1). This allowed us to identify i) elevated VOC biosynthetic activity in the appendix of *A. maculatum* during anthesis (Figure 2), ii) candidate transcripts for the production of several VOCs such as p-cresol, 2-heptanone, and sesquiterpenes (Figures 4 and 5), and iii) covariation between male floret terpene synthase expression and the relative proportions of *Psychoda phalaenoides* and *P. grisescens* trapped within inflorescences (Appendix S1, Figure S6).

5.1 – Differential expression analyses

During anthesis, diverse VOC metabolic processes were significantly enriched in *A. maculatum* appendix tissue when compared against male floret tissue. This result is consistent with previous research, suggesting that the thermogenic appendix plays a central role in diffusing

the potent VOC bouquets of Araceae (Meeuse 1975; Angioy *et al.* 2004; Barthlott *et al.* 2009). Although we expected to observe elevated TPS activity in male florets during anthesis compared to the morning following anthesis, it appears that many terpene synthases continued to be expressed at approximately equal levels in paired control samples collected the following morning; several factors may explain this pattern. First, given that sesquiterpenes are known to be synthesized and stored in *A. italicum* male floret tissue during the days leading up to anthesis (Leguet *et al.* 2014), TPS expression may have peaked prior to our during-anthesis tissue sampling. However, Araceae are known to emit particularly concentrated VOC blends during anthesis (Skubatz *et al.* 1995, 1996), which should require continuous biosynthesis and emission (Widhalm *et al.* 2015). It is therefore likely that the large number of terpene synthases we identified in the *A. maculatum* transcriptome include those which are most relevant to pollinator attraction.

The timing of our post-anthesis tissue sampling may have also influenced the above result. Post-anthesis sampling occurred within one hour after we collected and preserved all pollinators within the floral chamber; it is therefore likely that most pollinators normally would have still been trapped at this time. Both early (Dormer 1960; Prime 1960; Lack and Diaz 1991) and more recent studies (Bröderbauer *et al.* 2013) of *A. maculatum* have highlighted morphological features related to pollinator retention, namely the sterile flowers partially blocking the exit of the trap, and downward-pointing papillate cells on the inner wall of the spathe. However, continued expression of some TPS transcripts during the morning following anthesis may also highlight sesquiterpenes putatively involved in pollinator retention within the trap chamber. Consequently, our paired ‘post-anthesis’ samples may not have acted as a negative control for comparison to during-anthesis samples, as we initially intended. Further research is needed to confirm whether the end of male floret VOC emissions may also serve as

a cue for pollinator release, which occurs during the afternoon on the second day of the pollination cycle (Prime 1960; Lack and Diaz 1991; Chartier *et al.* 2013).

After comparing male floret transcript expression among populations during anthesis, we did not identify any putative VOC synthases which were unique to Forêt du Gâvre, France. High inter-individual variation in VOC emissions has been widely observed in *A. maculatum* (Szenteczki *et al.* 2021; Gfrerer *et al.* 2021) and this variability is also evident in our transcriptomic dataset. Although our differential expression analyses were unable to identify unique VOC synthases in the Forêt du Gâvre population, targeted investigations into candidate genes underlying key compound classes did provide further insights into species-specific pollinator attraction in *A. maculatum*.

5.2 – Candidate transcripts linked to *A. maculatum* VOCs

Most inter-individual variation in *A. maculatum* floral scent centres around variation in the ratio of aromatic amino acid-derived VOCs to sesquiterpenes (Szenteczki *et al.* 2021). Correspondingly, we identified abundant expression of transcripts putatively involved in the biosynthesis of these two compound classes. As predicted, indole synthesis via the tryptophan synthase alpha subunit (TSA) was highly expressed in appendix tissue in all studied populations. Some of the TSA transcripts we identified also appear to be homologs of BX1, which more efficiently cleaves indole-3-glycerol phosphate without any interaction with the beta subunit (Frey *et al.* 1997; Kriechbaumer *et al.* 2008). Conversely, TSB was more abundantly expressed in male floret tissue, which would limit indole emissions within the floral chamber.

Skatole (3-methylindole) has also been reported in *A. maculatum* VOC emissions (Szenteczki *et al.* 2021; Gfrerer *et al.* 2021), but we did not identify any homologs of indoleacetate decarboxylase in the *A. maculatum* transcriptome, which would produce this compound (R. *et al.* 2008). This may be due to low emissions of skatole by inflorescences sampled in this study;

ultimately, the mechanism by which *A. maculatum* produces skatole remains unclear. Prior to this study, it was also unclear whether *A. maculatum* produced p-cresol (4-methylphenol) via a pathway homologous to tyrosine degradation in bacteria (Selmer and Andrei 2001; Saito *et al.* 2018), or using one or more novel proteins. We were able to identify candidate transcripts (*i.e.* ISS1 and HPPR homologs) which may catalyse the conversion of tyrosine to p-coumaric acid, but were ultimately unable to identify transcripts which could further produce p-cresol from p-coumaric acid.

Homologs of phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H), which produce p-coumaric acid from phenylalanine, were also abundantly expressed in the appendix during anthesis. However, downstream proteins producing common floral benzenoid/phenylpropanoid (FBP) VOCs (e.g. eugenol, isoeugenol, benzyl benzoate, benzaldehyde, and phenylacetaldehyde (Muhlemann *et al.* 2014)) were largely absent or lowly expressed in both our VOC and transcriptomic datasets. Typically, p-coumarate-3-hydroxylase (C3H) would further modify p-coumaric acid into precursors for the aforementioned FBP VOCs, but this gene was expressed at relatively low levels in appendix tissue. Interestingly, experimental knock-outs of C3H in *Petunia × hybrida* have demonstrated that downregulation of this gene also leads to unexpected production of p-cresol (Kim *et al.* 2019). Given that we observed similar patterns in *A. maculatum*, it appears that flowering plants may have a novel mechanism for producing p-cresol involving p-coumaric acid, which is distinct from known bacterial proteins such as hydroxyphenylacetate decarboxylase.

While not derived from amino acids, 2-heptanone is another abundant *A. maculatum* VOC with a ‘fermented’ scent. In this study, we were able to identify several methylketone synthase homologs (MKS1 and MKS2), which produce 2-heptanone and related compounds in tomato plants (Ben-Israel *et al.* 2009; Khuat *et al.* 2019). 2-heptanone is known to be attractive to fruit flies (Prokopy *et al.* 1998) and beetles (Torto *et al.* 2007); however, it has not yet been linked

with the attraction of specific pollinators of *A. maculatum* (Espíndola *et al.* 2011; Szenteczki *et al.* 2021). While this compound is infrequent in VOC bouquets, it can be a proportionally large component of VOC emissions when it is produced (Szenteczki *et al.* 2021); further research is therefore needed to determine whether this compound is also under selection (*e.g.* by *Drosophilidae*).

Finally, we identified many putative terpene synthases in the *A. maculatum* transcriptome, the majority of which appear to produce sesquiterpene VOCs. These include cytochrome P450 enzymes (P450s), which can further modify volatile terpenes through oxidation, methylation, or acylation (Hamberger and Bak 2013). P450s were the first proteins identified in *Arum* appendices (Yáhiel *et al.* 1974), and our transcriptomic data further confirm that a diverse suite of P450s are expressed in *A. maculatum* appendix and male floret tissue during anthesis. Notably, we found that a trimethyltridecatetraene synthase (Cyt P450 92C6) homolog appears to be correlated with the production of at least one unnamed *A. maculatum* sesquiterpene (Kovats RI 1681). This unnamed sesquiterpene was the single strongest predictor of *P. grisescens* attraction in a large-scale survey of *A. maculatum* pollinators (Szenteczki *et al.* 2021). Until now, it has not been possible to experimentally test whether this compound alone is attractive to *Psychodidae*. However, the candidate gene we identified could be useful in future research aiming to produce this compound through heterologous expression in bacteria (Komatsu *et al.* 2010, 2013).

Our results also confirm that bicyclogermacrene synthase (Crocoll *et al.* 2010) is consistently and almost exclusively expressed in male floret tissue during anthesis, in accordance with the dominance of this sesquiterpene in *Arum maculatum* floral chamber scent (Kite *et al.* 1998). Interestingly, previous studies have demonstrated that 9-methyl germacrene B is a sex pheromone produced by male *Lutzomyia longipalpis* (*Psychodidae*) to attract females (Hamilton *et al.* 1996; Gordon C. Hamilton *et al.* 1999). Consequently, *A. maculatum* may emit

bicyclogermacrene, or a closely related sesquiterpene compound as a as part of their deceptive pollination strategy (*i.e.* as a pheromone mimic). Terpene emissions in the floral chamber could also stimulate the movement of trapped Psychodidae over the male and female florets, which could aid in pollination and pollen dispersal; however, further experiments are needed to test these hypotheses. Gas chromatography–electroantennography (GC–EAD; (Schiestl and Marion-Poll 2002)) would allow for more precise identification of bioactive compounds that elicit sex- and species-specific reactions in the antennae of Psychodidae, and the candidate terpene synthases identified in this study may be useful in guiding such research.

5.3 – Male floret-specific terpene synthases and their influence on pollinator attraction

There is growing evidence to suggest that tissue-specific transcript expression of VOC synthases is a common characteristic of deceptive pollination systems ((Wong *et al.* 2017a), and references therein), and our results confirm similar patterns in *A. maculatum*. Specifically, our coinertia analyses identified significant covariation between male floret terpene synthase expression, and the communities of Psychodidae pollinators trapped by inflorescences. While the floral scent of *A. maculatum* is often described as dung-like due to abundant emissions of amino acid-derived VOCs, our results suggest that sesquiterpene VOCs may be subject to pollinator-mediated selection as well. Recent large-scale ecological studies also support this hypothesis, with sesquiterpene compounds, rather than so-called “dung-mimicking” compounds, being a better predictor of variation in pollinator attraction patterns (Szenteczki *et al.* 2021) and fruit set size (Gfrerer *et al.* 2021) in *A. maculatum*.

Some of the main compounds related to sex- and species-specific attraction in our coinertia analysis included humulene (covarying with female *P. phalaenoides*), and the aforementioned trimethyltridecatetraene synthase homolog, which may produce an unnamed sesquiterpene with a Kovats Retention Index of 1681 (covarying with *P. grisescens*). Notably, these same compounds were highlighted as key predictors of pollinator species trapped by *A. maculatum*

inflorescences in a random forest analysis with a larger sample size (Szenteczki *et al.* 2021). Further research is therefore needed in order to assess the importance of male floret VOCs for pollinator attraction in *A. maculatum*, and to confirm the true identities of the compounds produced by the candidate terpene synthases we identified.

6. CONCLUSION

It appears that *A. maculatum* inflorescences employ a combination of highly diverse appendix VOCs, including 2-heptanone, indole, and p-cresol to attract a broad range of coprophilous dipterans to their inflorescences, and specialized sesquiterpene emissions to further lure specific Psychodidae into the floral trap chamber. Male floret-specific VOCs, particularly bicyclogermacrene, may also play a role in retaining pollinators until the pollination cycle is complete. To our knowledge, no studies have specifically investigated trap chamber VOC profiles since Kite and colleagues (Kite 1995; Kite *et al.* 1998) first characterized *A. maculatum* floral scent using GC-MS. Taken together, our transcriptomic data indicate that volatile sesquiterpene terpene emissions in the floral trap chamber are an important aspect of this highly specialized lure-and-trap pollination system, which has been overlooked for decades.

ACKNOWLEDGEMENTS

We thank Monica Fleisher, Andrea Galmán, and Alberto Garcia Jimenez for their dedicated assistance during our field sampling, and Laurent Oppliger and colleagues for allowing us to sample VOCs and tissue from the native population of *A. maculatum* at the Jardin Botanique de Neuchâtel. We also thank Dessislava Savova Bianchi for her support with RNA extractions, Johann Weber for his advice on our RNA-seq library preparation, and Eran Pichersky for his advice on the biosynthesis of some floral VOCs.

FUNDING STATEMENT

This project was funded by the Swiss National Science Foundation through grant P1NEP3_191659 awarded to MAS and grant 31003A_163334 awarded to NA and SR.

DATA ACCESSIBILITY

Appendices S1 and S2 contains supporting figures and tables referenced in the main text. Raw Illumina RNA-seq reads have been deposited in the NCBI SRA database under the BioProject ID #TBD (<https://www.ncbi.nlm.nih.gov/bioproject/#TBD>). Additional data, including the files and R code needed to reproduce all analyses and figures have been archived in the Zenodo repository #TBD (<https://dx.doi.org/zenodo.#TBD>).

COMPETING INTERESTS

The authors declare no competing interests. The above-mentioned funding agencies did not have any influence on study design, interpretation of results, or decision to publish.

AUTHORS' CONTRIBUTIONS

M.A.S., A.L.G., M.G., S.R., and N.A. designed the experiments. M.A.S. and A.L.G. conducted the field sampling. M.A.S. prepared tissue samples for RNA-Seq library preparation, and A.L.G. and M.G. analysed the headspace VOC samples and identified pollinators. M.A.S. and J.G. designed and performed the bioinformatic analyses, and M.G. contributed to the interpretation of the results. M.A.S. wrote the manuscript. All co-authors contributed to the revision of the manuscript, and have approved it for publication

REFERENCES

- Alexa, A., and J. Rahnenfuhrer, 2010 topGO: enrichment analysis for gene ontology. R package version 2.0.
- Angioy, A. M., M. C. Stensmyr, I. Urru, M. Puliafito, I. Collu *et al.*, 2004 Function of the heater: the dead horse arum revisited. *Proceedings. Biol. Sci.* 271 Suppl: S13-5.
- Barthlott, W., J. Szarzynski, P. Vlek, W. Lobin, and N. Korotkova, 2009 A torch in the rain forest: thermogenesis of the Titan arum (*Amorphophallus titanum*). *Plant Biol. (Stuttg.)* 11: 499–505.
- Ben-Israel, I., G. Yu, M. B. Austin, N. Bhuiyan, M. Auldridge *et al.*, 2009 Multiple biochemical and morphological factors underlie the production of methylketones in tomato trichomes. *Plant Physiol.* 151: 1952–1964.
- Bröderbauer, D., A. Weber, and A. Diaz, 2013 The design of trapping devices in pollination traps of the genus *Arum* (Araceae) is related to insect type. *Bot. J. Linn. Soc.* 172: 385–397.
- Byers, K. J. R. P., H. D. Bradshaw, and J. A. Riffell, 2014 Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (*Mimulus*). *J. Exp. Biol.* 217: 614–623.
- Chartier, M., L. Pélozuelo, B. Buatois, J.-M. Bessière, and M. Gibernau, 2013 Geographical variations of odour and pollinators, and test for local adaptation by reciprocal transplant of two European *Arum* species. *Funct. Ecol.* 27: 1367–1381.
- Chartier, M., L. Pélozuelo, and M. Gibernau, 2011 Do floral odor profiles geographically vary with the degree of specificity for pollinators? Investigation in two sapromyophilous *Arum* species (Araceae). *Ann. la Société Entomol. Fr.* 47: 71–77.
- Chen, S., Y. Zhou, Y. Chen, and J. Gu, 2018 fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34: i884–i890.
- Crocoll, C., J. Asbach, J. Novak, J. Gershenson, and J. Degenhardt, 2010 Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. *Plant Mol. Biol.* 73: 587–603.
- Dhandapani, S., J. Jin, V. Sridhar, R. Sarojam, N.-H. Chua *et al.*, 2017 Integrated metabolome and transcriptome analysis of *Magnolia champaca* identifies biosynthetic pathways for floral volatile organic compounds. *BMC Genomics* 18: 463.
- Diaz, A., and G. Kite, 2002 A comparison of the pollination ecology of *Arum maculatum* and *A. italicum* in England. *Watsonia* 24: 171–182.
- Dobson, H. E. M., J. Arroyo, G. Bergström, and I. Groth, 1997 Interspecific variation in floral fragrances within the genus *Narcissus* (Amaryllidaceae). *Biochem. Syst. Ecol.* 25: 685–706.
- Dormer, K. J., 1960 The truth about pollination in *Arum*. *New Phytol.* 59: 298–301.
- Espíndola, A., and N. Alvarez, 2011 Comparative Phylogeography in a Specific and Obligate

Pollination Antagonism. PLoS One 6: e28662.

- Espíndola, A., L. Pellissier, and N. Alvarez, 2011 Variation in the proportion of flower visitors of *Arum maculatum* along its distributional range in relation with community-based climatic niche analyses. *Oikos* 120: 728–734.
- Fenster, C. B., W. S. Armbruster, P. Wilson, M. R. Dudash, and J. D. Thomson, 2004 Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.* 35: 375–403.
- Frey, M., P. Chomet, E. Glawischnig, C. Stettner, S. Grün *et al.*, 1997 Analysis of a chemical plant defense mechanism in grasses. *Science* 277: 696–699.
- Friberg, M., C. Schwind, P. R. Guimarães, R. A. Raguso, and J. N. Thompson, 2019 Extreme diversification of floral volatiles within and among species of *Lithophragma* (Saxifragaceae). *Proc. Natl. Acad. Sci.* 116: 4406–4415.
- Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li, 2012 CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152.
- Gervasi, D. D. L., and F. P. Schiestl, 2017 Real-time divergent evolution in plants driven by pollinators. *Nat. Commun.* 8: 14691.
- Gfrerer, E., D. Laina, M. Gibernau, R. Fuchs, M. Happ *et al.*, 2021 Floral Scents of a Deceptive Plant Are Hyperdiverse and Under Population-Specific Phenotypic Selection. *Front. Plant Sci.* 12: 719092.
- Gibernau, M., 2004 Pollination in the genus *Arum* – a review. *Aroideana* 27: 148–166.
- Gilbert, D. G., 2019 Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? *bioRxiv* 829184.
- Gordon C. Hamilton, J., H. C. Ibbotson, A. M. Hooper, and J. A. Pickett, 1999 9-Methylgermacrene-B is confirmed as the sex pheromone of the sandfly *Lutzomyia longipalpis* from Lapinha, Brazil, and the absolute stereochemistry defined as S. *Chem. Commun.* 2335–2336.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29: 644–652.
- Hamberger, B., and S. Bak, 2013 Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* 368: 20120426.
- Hamilton, J. G., G. W. Dawson, and J. A. Pickett, 1996 9-Methylgermacrene-B; proposed structure for novel homosesquiterpene from the sex pheromone glands of *Lutzomyia longipalpis* (Diptera: Psychodidae) from Lapinha, Brazil. *J. Chem. Ecol.* 22: 1477–1491.
- Hernández-Hernández, T., and J. J. Wiens, 2020 Why Are There So Many Flowering Plants? A Multiscale Analysis of Plant Diversification. *Am. Nat.* 195: 948–963.

- Hervé, M., 2021 RVAideMemoire: Diverse basic statistical and graphical functions. R package version 0.9-81.
- Holopainen, J. K., 2004 Multiple functions of inducible plant volatiles. *Trends Plant Sci.* 9: 529–533.
- Irwin, R. E., L. S. Adler, and A. K. Brody, 2004 The dual role of floral traits: Pollinator attraction and plant defense. *Ecology* 85: 1503–1511.
- Ježek, J., 1990 Redescriptions of nine common palaeartic and holarctic species of Psychodini end. (Diptera: Psychodidae). *Acta Entomol. Musei Natl. Pragae* 43: 33–83.
- Junker, R. R., and A. L. Parachnowitsch, 2015 Working towards a holistic view on flower traits - how floral scents mediate plant-animal interactions in concert with other floral characters. *J. Indian Inst. Sci.* 95: 43–67.
- Khuat, V. L. U., V. T. T. Bui, H. T. D. Tran, N. X. Truong, T. C. Nguyen *et al.*, 2019 Characterization of *Solanum melongena* Thioesterases Related to Tomato Methylketone Synthase 2. *Genes (Basel)*. 10: 549.
- Kim, J. Y., R. T. Swanson, M. I. Alvarez, T. S. Johnson, K. H. Cho *et al.*, 2019 Down regulation of p-coumarate 3-hydroxylase in petunia uniquely alters the profile of emitted floral volatiles. *Sci. Rep.* 9: 8852.
- Kite, G. C., 1995 The floral odour of *Arum maculatum*. *Biochem. Syst. Ecol.* 23: 343–354.
- Kite, G. C., W. L. A. Hetterscheid, M. J. Lewis, P. C. Boyce, J. Ollerton *et al.*, 1998 Inflorescence odours and pollinators of *Arum* and *Amorphophallus* (Araceae), pp. 295–315 in *Reproductive Biology*, edited by S. J. Owens and P. J. Rudall. Royal Botanic Gardens Kew, London, UK.
- Knoll, F., 1926 Insekten und Blumen, die *Arum* - Blütenstände und ihre Besucher. *Anhandlungen der Zool. und Bot. Gesellschaft Wien* 12: 379–481.
- Knudsen, J. T., and J. Gershenzon, 2020 The Chemical Diversity of Floral Scent, pp. 57–78 in *Biology of Plant Volatiles*, edited by E. Pichersky and N. Dudareva. CRC Press.
- Knudsen, J. T., and L. Tollsten, 1993 Trends in floral scent chemistry in pollination syndromes: floral scent composition in moth-pollinated taxa. *Bot. J. Linn. Soc.* 113: 263–284.
- Kolde, R., 2019 pheatmap: Pretty Heatmaps. R package version 1.0.12.
- Komatsu, M., K. Komatsu, H. Koiwai, Y. Yamada, I. Kozono *et al.*, 2013 Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* 2: 384–396.
- Komatsu, M., T. Uchiyama, S. Omura, D. E. Cane, and H. Ikeda, 2010 Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 107: 2646–2651.
- Kriechbaumer, V., L. Weigang, A. Fiesselmann, T. Letzel, M. Frey *et al.*, 2008 Characterisation of the tryptophan synthase alpha subunit in maize. *BMC Plant Biol.* 8:

44.

- Lack, A. J., and A. Diaz, 1991 The pollination of *Arum maculatum* L. - a historical review and new observations. *Watsonia* 18: 333–342.
- Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.
- Leguet, A., M. Gibernau, L. Shintu, S. Caldarelli, S. Moja *et al.*, 2014 Evidence for early intracellular accumulation of volatile compounds during spadix development in *Arum italicum* L. and preliminary data on some tropical Aroids. *Naturwissenschaften* 101: 623–635.
- Letunic, I., and P. Bork, 2021 Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49: W293–W296.
- Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323.
- Li, W., and A. Godzik, 2006 Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659.
- Liu, D., Y. Wei, X. Liu, Y. Zhou, L. Jiang *et al.*, 2018 Indoleacetate decarboxylase is a glyceryl radical enzyme catalysing the formation of malodorant skatole. *Nat. Commun.* 9: 4224.
- Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15: 550.
- Meeuse, B. J. D., 1975 Thermogenic Respiration in Aroids. *Annu. Rev. Plant Physiol.* 26: 117–126.
- Moriya, Y., M. Itoh, S. Okuda, A. C. Yoshizawa, and M. Kanehisa, 2007 KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35: 182–185.
- Muhlemann, J. K., A. Klempien, and N. Dudareva, 2014 Floral volatiles: from biosynthesis to function. *Plant. Cell Environ.* 37: 1936–1949.
- Newman, E., B. Anderson, and S. D. Johnson, 2012 Flower colour adaptation in a mimetic orchid. *Proceedings. Biol. Sci.* 279: 2309–2313.
- Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh, 2015 IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* 32: 268–274.
- van der Niet, T., and S. D. Johnson, 2012 Phylogenetic evidence for pollinator-driven diversification of angiosperms. *Trends Ecol. Evol.* 27: 353–361.
- Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre *et al.*, 2020 vegan: Community Ecology Package. R package version 2.5-7.
- Onda, Y., K. Mochida, T. Yoshida, T. Sakurai, R. S. Seymour *et al.*, 2015 Transcriptome analysis of thermogenic *Arum concinatum* reveals the molecular components of floral

- scent production. *Sci. Rep.* 5:.
- Paiva, É. A. S., I. Ballego-Campos, and M. Gibernau, 2021 True nectar or stigmatic secretion? Structural evidence elucidates an old controversy regarding nectaries in *Anthurium*. *Am. J. Bot.* 108: 37–50.
- Pellmyr, O., and L. B. Thien, 1986 Insect Reproduction and Floral Fragrances : Keys to the Evolution of the Angiosperms? *Taxon* 35: 76–85.
- Pichersky, E., and J. Gershenzon, 2002 The formation and function of plant volatiles: Perfumes for pollinator attraction and defense. *Curr. Opin. Plant Biol.* 5: 237–243.
- Prime, C. T., 1960 *Lords and ladies*. Collins, London.
- Prokopy, R. J., X. Hu, E. B. Jang, R. I. Vargas, and J. D. Warthen, 1998 Attraction of mature *Ceratitidis capitata* females to 2-heptanone, a component of coffee fruit odor. *J. Chem. Ecol.* 24: 1293–1304.
- R., W. T., P. N. P., D. H. L., and C. M. A., 2008 Catabolic Pathway for the Production of Skatole and Indoleacetic Acid by the Acetogen *Clostridium drakei*, *Clostridium scatologenes*, and Swine Manure. *Appl. Environ. Microbiol.* 74: 1950–1953.
- R Core Team, 2021 R: A Language and Environment for Statistical Computing.
- Saito, Y., T. Sato, K. Nomoto, and H. Tsuji, 2018 Identification of phenol- and p-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiol. Ecol.* 94:.
- Sayols, S., 2020 rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms. R package version 1.4.
- Schemske, D. W., and H. D. Bradshaw, 1999 Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc. Natl. Acad. Sci.* 96: 11910 LP – 11915.
- Schiestl, F. P., and S. D. Johnson, 2013 Pollinator-mediated evolution of floral signals. *Trends Ecol. Evol.* 28: 307–315.
- Schiestl, F. P., and F. Marion-Poll, 2002 Detection of physiologically active flower volatiles using gas chromatography coupled with electroantennography, pp. 173–198 in *Analysis of taste and aroma*, Springer.
- Schmucker, T., 1925 Beiträge zur Biologie und Physiologie von *Arum maculatum*. *Flora oder Allg. Bot. Zeitung* 460–475.
- Scott, C., T. Pierce, L. K. Johnson, C. T. Brown, C. Reid *et al.*, 2019 dib-lab/dammit: dammit v1.2.
- Selmer, T., and P. I. Andrei, 2001 p-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glycyl radical enzyme catalysing the formation of p-cresol. *Eur. J. Biochem.* 268: 1363–1372.
- Sepey, M., M. Manni, and E. M. Zdobnov, 2019 BUSCO: Assessing Genome Assembly and

- Annotation Completeness BT - Gene Prediction: Methods and Protocols, pp. 227–245 in edited by M. Kollmar. Springer New York, New York, NY.
- Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus *et al.*, 2011 Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7: 539.
- Skubatz, H., D. D. Kunkel, W. N. Howald, R. Trenkle, and B. Mookherjee, 1996 The *Sauromatum guttatum* appendix as an osmophore: Excretory pathways, composition of volatiles and attractiveness to insects. *New Phytol.* 134: 631–640.
- Skubatz, H., D. D. Kunkel, J. M. Patt, W. N. Howald, T. G. Hartman *et al.*, 1995 Pathway of terpene excretion by the appendix of *Sauromatum guttatum*. *Proc. Natl. Acad. Sci. U. S. A.* 92: 10084–10088.
- Soneson, C., M. I. Love, and M. D. Robinson, 2015 Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4: 1521.
- Song, L., and L. Florea, 2015 Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* 4: 48.
- Supek, F., M. Bošnjak, N. Škunca, and T. Šmuc, 2011 REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6: e21800.
- Szentezcki, M. A., A. L. Godschalx, A. Galmán, A. Espíndola, M. Gibernau *et al.*, 2021 Spatial and temporal heterogeneity in pollinator communities maintains within-species floral odour variation. *Oikos* 130: 1487–1499.
- Thioulouse, J., S. Dray, A.-B. Dufour, A. Siberchicot, T. Jombart *et al.*, 2018 *Multivariate Analysis of Ecological Data with ade4*. Springer, New York, NY.
- Thompson, J. N., C. Schwind, and M. Friberg, 2017 Diversification of Trait Combinations in Coevolving Plant and Insect Lineages. *Am. Nat.* 190: 171–184.
- Thompson, J. N., C. Schwind, P. R. Guimarães, and M. Friberg, 2013 Diversification through multitrait evolution in a coevolving interaction. *Proc. Natl. Acad. Sci.* 110: 11487–11492.
- Torto, B., D. G. Boucias, R. T. Arbogast, J. H. Tumlinson, and P. E. A. Teal, 2007 Multitrophic interaction facilitates parasite–host relationship between an invasive beetle and the honey bee. *Proc. Natl. Acad. Sci.* 104: 8374 LP – 8378.
- Trunschke, J., K. Lunau, G. H. Pyke, Z.-X. Ren, and H. Wang, 2021 Flower Color Evolution and the Evidence of Pollinator-Mediated Selection. *Front. Plant Sci.* 12: 1096.
- Urru, I., J. Stökl, J. Linz, T. Krügel, M. C. Stensmyr *et al.*, 2010 Pollination strategies in Cretan Arum lilies. *Biol. J. Linn. Soc.* 101: 991–1001.
- Vogel, S., 1965 Kesselfallen–Blumen. *Umschau* 65: 12–17.
- Whitehead, M. R., and R. Peakall, 2009 Integrating floral scent, pollination ecology and population genetics. *Funct. Ecol.* 23: 863–874.

- Widhalm, J. R., R. Jaini, J. A. Morgan, and N. Dudareva, 2015 Rethinking how volatiles are released from plant cells. *Trends Plant Sci.* 20: 545–550.
- Wong, D. C. J., R. Amarasinghe, C. Rodriguez-Delgado, R. Eyles, E. Pichersky *et al.*, 2017a Tissue-Specific Floral Transcriptome Analysis of the Sexually Deceptive Orchid *Chiloglottis trapeziformis* Provides Insights into the Biosynthesis and Regulation of Its Unique UV-B Dependent Floral Volatile, Chiloglottone 1. *Front. Plant Sci.* 8: 1260.
- Wong, D. C. J., E. Pichersky, and R. Peakall, 2017b The biosynthesis of unusual floral volatiles and blends involved in orchid pollination by deception: Current progress and future prospects. *Front. Plant Sci.* 8: 1955.
- Xiao, H., Y. Zhang, and M. Wang, 2019 Discovery and Engineering of Cytochrome P450s for Terpenoid Biosynthesis. *Trends Biotechnol.* 37: 618–631.
- Yáhiel, V., M. G. Cotte-Martinon, and G. Ducet, 1974 Un cytochrome de type P450 dans le spadice d'arum. *Phytochemistry* 13: 1649–1651.
- Zhu, A., J. G. Ibrahim, and M. I. Love, 2019 Heavy-Tailed prior distributions for sequence count data: Removing the noise and preserving large differences. *Bioinformatics* 35: 2084–2092.

Figure 1. Overview of the sampling design, bioinformatic pipeline, and differential expression analyses in this study. Inset photo: *Arum maculatum* inflorescence, with the lower spathe chamber dissected to reveal the male and female florets. Inset map: population codes, placed over their approximate locations.

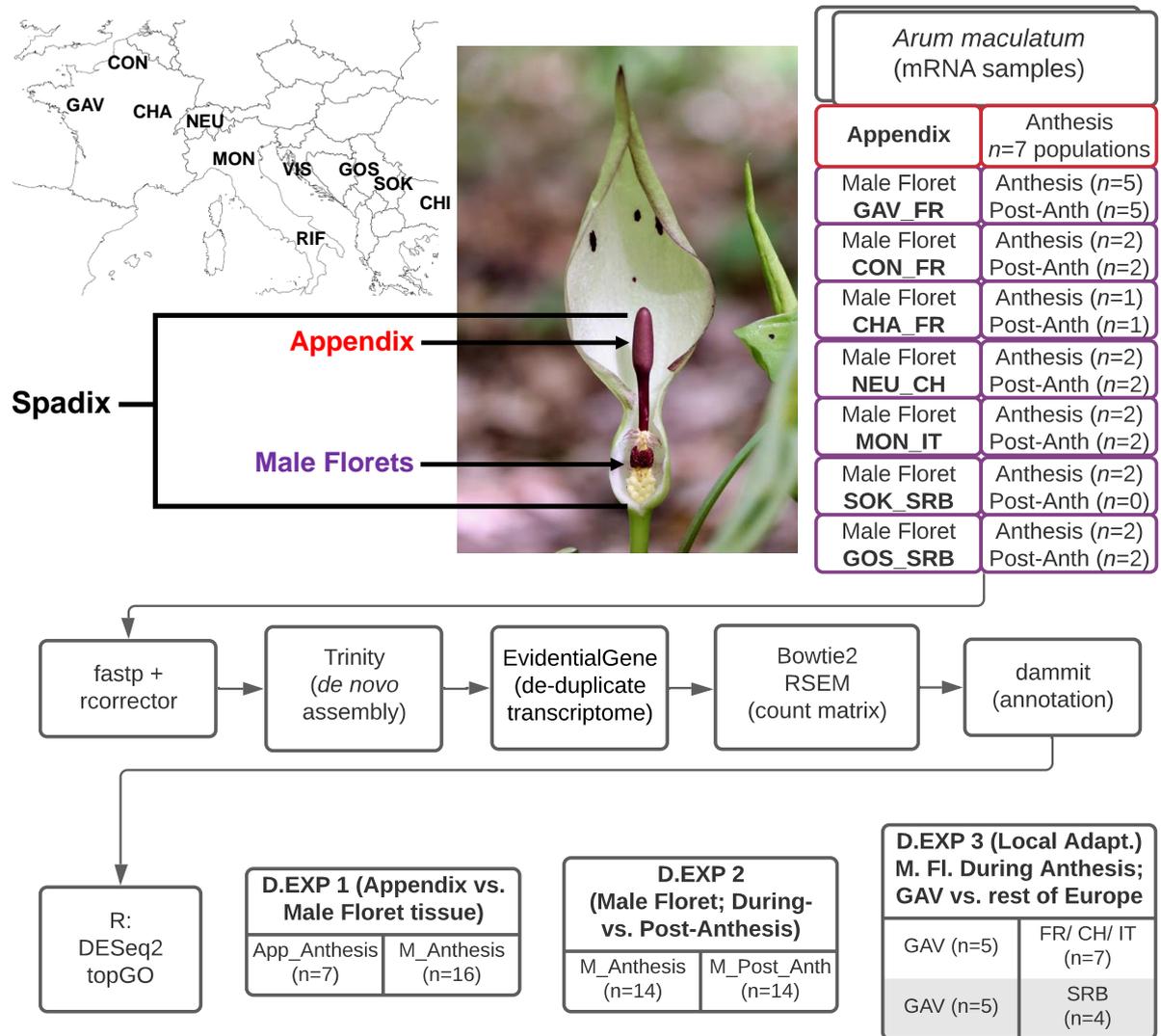


Figure 3. Significantly enriched GO terms when comparing *Arum maculatum* transcript expression in male floret tissue during anthesis ($n = 14$), versus approximately 18 hours after anthesis ($n = 14$). Non-redundant GO terms are visualized in semantic similarity space (allowed similarity = 0.8); the full list of GO terms represented above is available in Appendix S2. [*] indicates parent groups linked to the biosynthesis of volatile compounds.



Figure 4. A simplified overview of key proteins involved in producing aromatic amino-acid derived *Arum maculatum* VOCs, and a heatmap of their *vst*-transformed expression during anthesis. Colour scale represents whether expression in a given group (*i.e.* tissue type and population) is above or below the transcript's mean expression across all samples. Genes in grey are absent from the *A. maculatum* transcriptome. Gene names: BEBT = benzyl alcohol O-benzoyltransferase; BSMT = benzoic acid/salicylic acid carboxyl methyltransferase; C4H = cinnamate 4-hydroxylase; DHQ = 3-dehydroquinate dehydratase / shikimate dehydrogenase; EGS = eugenol synthase; HPPR = hydroxyphenylpyruvate reductase; ISS1 = aromatic aminotransferase; MKS = methylketone synthase; PAAS = phenylacetaldehyde synthase; PAL = phenylalanine ammonialyase; TSA/B = tryptophan synthase alpha/beta subunit; ? = unknown gene(s).

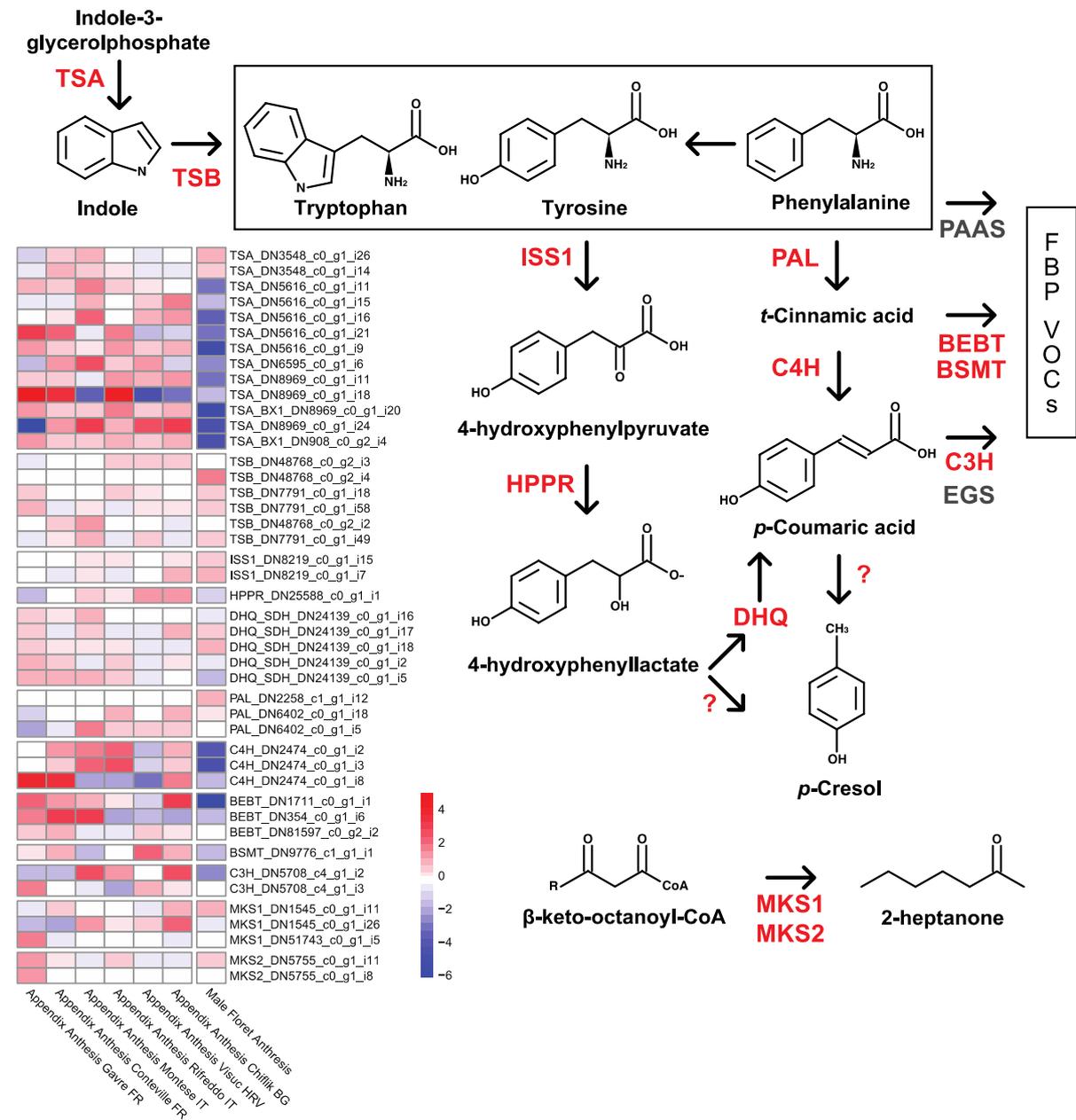


Figure 5. Heatmap of *vst*-transformed expression of putative terpene synthases in *Arum maculatum*. Colour scale represents whether expression in a given group (*i.e.* tissue type, stage of anthesis, and population) is above or below the transcript's mean expression across all samples.

