Identification and characterisation of *Planococcus citri cis*- and *trans*-isoprenyl diphosphate synthase genes, supported by short- and long-read transcriptome data

Mojca Juteršek[1,2,#,*], Iryna M. Gerasymenko[3,4,#], Marko Petek[1], Elisabeth Haumann[3,4], Sandra Vacas[5], Kalyani Kallam[6], Silvia Gianoglio[7], Vicente Navarro-Llopis[5], Ismael Navarro Fuertes[8], Nicola Patron[6], Diego Orzáez[7], Kristina Gruden[1], Heribert Warzecha[3,4], Špela Baebler[1]

[1] National Institute of Biology, Department of Biotechnology and Systems Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia

[2] Jožef Stefan International Postgraduate School, Jamova 39, SI-1000 Ljubljana, Slovenia

[3] Plant Biotechnology and Metabolic Engineering, Department of Biology, Technical University of Darmstadt, Schnittspahnstrasse 4, 64287 Darmstadt, Germany

[4] Centre for Synthetic Biology, Technical University of Darmstadt, Schnittspahnstrasse 4, 64287 Darmstadt, Germany

[5] Instituto Agroforestal del Mediterráneo-CEQA, Universitat Politècnica de València. Camino de Vera s/n, Valencia, Spain

[6] Engineering Biology, Earlham Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UZ, UK

[7] Institute for Plant Molecular and Cell Biology (IBMCP), Consejo Superior de Investigaciones Científicas (CSIC) - Universitat Politècnica de València (UPV), Valencia, Spain

[8] Department of Organic Chemistry, University of Valencia, Burjassot, València, Spain

[#] These authors contributed equally to the work.

*Coresspondence to mojca.jutersek@nib.si

Abstract

Many insect species rely on diverse terpenoids for their development and interorganismal interactions. However, little is known about terpenoid biosynthesis in insects. The monoterpenoid sex pheromones of mealybugs and scale insects (Coccoidea) are particularly enigmatic, with several species producing unique structures presumed to result from the irregular coupling activity of unidentified isopentenyl diphosphate synthases (IDSs). Enzymes capable of similar transformations have previously only been described from a few plant, bacterial and archaeal species. To investigate if insect irregular monoterpenes can be biosynthesised by similar enzymes, we performed a comprehensive search for IDS coding sequences in the genome of *Planococcus citri*, a widespread agricultural pest. We complemented the available *P. citri* genome data with newly generated short- and long-read transcriptome data. The identified candidate genes had homology to both short- and long-chain IDSs and some appeared to be paralogous, indicating gene duplications and consequent IDS gene family expansion in *P. citri*. We tested the activity of eleven candidate gene products, confirming *in vitro* regular activity for five enzymes, one of which (*trans*IDS5) also produced the irregular prenyl diphosphates, maconelliyl and lavandulyl diphosphate. Targeted mutagenesis of selected aspartates and a lysine in the active site of *trans*IDS5 uncovered their importance for chain-length preference and irregular coupling. This work provides an important foundation for deciphering terpenoid biosynthesis in mealybugs, as well as a potential source of enzymes for the biotechnological production of sustainable insect pest management products.

Keywords

Planococcus citri, isopentenyl diphosphate synthase, transcriptome assembly, Iso-Seq, irregular coupling, maconellily diphosphate, lavadulyl diphosphate, planococcyl diphosphate, insect sex pheromones

54 ## Introduction

55 Terpenoids are a large class of metabolites produced by organisms in every branch of life. In
56 insects, terpenoids have important roles in development (e.g., juvenile and moulting hormones)
57 and as semiochemicals in intra- as well as interspecific interactions (e.g., sex, aggregating,
58 alarm, dispersal, maturation, anti-aphrodisiac or trail pheromones, and defence compounds)
59 (Tillman et al., 1999; Beran et al., 2019). Of special interest are the terpenoid sex pheromones
60 produced by some mealybug and scale insect species (order Hemiptera, Superfamily
61 Coccoidea). Sex pheromones are species-specific, volatile, long-range attractants. In
62 Coccoidea, they are produced by females to facilitate male navigation during mating and are
63 thus emitted by virgin females, with cessation of production after mating (Zou and Millar,
64 2015). The majority of mealybug (Family Pseudococcidae) and armoured scale insect (Family
65 Diaspididae) sex pheromones identified to date are irregular terpenoid compounds, mostly
66 esterified mono- or sesquiterpenoid alcohols and carboxylic acids (Zou and Millar, 2015;
67 Franco et al., 2022). Chemically synthesised mealybug and scale insect pheromones have been
68 successfully used for pest control, providing a sustainable measure in climate
69 change-challenged agriculture (Vacas et al., 2010; Zou et al., 2013; Zou and Millar, 2015;
70 Lucchi et al., 2019; Daane et al., 2021). Although an extensive body of research has been
71 dedicated to the identification and study of mealybug and scale insect sex pheromones, the
72 biosynthesis of their irregular backbone remains elusive.

73 The first step of terpenoid biosynthesis is catalysed by isoprenyl diphosphate synthases (IDSs
74 or prenyltransferases), which couple the C5 terpene building blocks, isopentenyl diphosphate
75 (IPP) and dimethylallyl diphosphate (DMAPP) into C10 (geranyl diphosphate, GPP), C15
76 (farnesyl diphosphate, FPP), C20 (geranylgeranyl diphosphate, GGPP) and longer prenyl
77 diphosphate precursor chains by sequential condensation steps (Nagel et al., 2019). The prenyl
78 diphosphate chains are then transformed into terpenoid compounds through the activity of
79 terpene synthases (TPSs).

80 IDSs can be classified as either *cis*- or *trans*-, depending on the stereochemistry of the double
81 bond in the reaction product. The two classes are evolutionarily and structurally distinct.
82 *Trans*-IDSs adopt an α-fold and contain two conserved aspartate-rich motifs known as FARM
83 (first aspartate rich motif) and SARM (second aspartate rich motif), which mainly occur as
84 "DDxx(x)D". *Cis*-IDSs, however, lack distinct conserved motifs and adopt the ζ-fold (Nagel
85 et al., 2019). Aspartate-rich motifs in *trans*-IDSs coordinate the Mg^{2+} ions important for

generation of the carbocation in the allylic substrate (DMAPP, GPP, FPP or other), which can be attacked by IPP to form the new carbon-carbon (C-C) bond in the alkylation reaction. The coupling step catalysed by IDSs can be also classified based on the orientation of the alkylation reaction. Most IDS enzymes catalyse the formation of regular 1'-4 (head-to-tail) C-C bond between the allylic substrate and IPP, while irregular head-to-middle reactions have been described for some *cis*- and *trans*-IDS enzymes, coupling two DMAPP units or DMAPP with an allylic substrate, forming branched or cyclic polyprenyl chains (Kobayashi and Kuzuyama, 2018). To date, only a few enzymes with irregular coupling activity have been identified from plants (Rivera et al., 2001; Demissie et al., 2013), bacteria (Ozaki et al., 2014), and archaea (Ogawa et al., 2016).

Planococcus citri (Risso, 1813), the citrus mealybug, is a small sucking insect. It is a widespread pest of numerous crops and ornamental plants, causing serious damage and economic losses (Watson, 2023). Its sex pheromone, planococcyl acetate, has an cyclobutane backbone, presumably resulting from irregular IDS coupling. It has been successfully chemically synthesised and deployed in integrated pest management via trap-based management strategies, which require only milligram quantities (Bierl-Leonhardt et al., 1981; Wolk et al., 1986; Dunkelblum et al., 2002; Passaro and Webster, 2004). Alternative management strategies, such as mating disruption, demand larger quantities and, therefore, require low-cost, scalable production methods (Zou and Millar, 2015). Biotechnological production of sex pheromones has been demonstrated recently for lepidopteran pheromones (Mateos-Fernández et al., 2021; Petkevicius et al., 2022) and biotechnological synthesis has been proposed as a sustainable route for large-scale, low-cost pheromone production. However, these methods require knowledge of the biosynthetic pathways.

To that end, we performed a comprehensive search of the *P. citri* genome, supported by newly generated short- and long-read sequence datasets of *P. citri* females, to identify candidate *cis*- and *trans*-IDSs. Eleven of the identified IDS sequences were tested for regular and irregular coupling activity, with a focus on the production of short-chain backbones. We confirmed regular coupling activity for five *trans*-IDS candidate enzymes, one of which also produced smaller amounts of the irregular lavandulyl and maconelliyl diphosphates. Our work contributes to understanding of mealybug terpenoid biosynthesis providing a foundation for its biotechnological exploitation.

117 # Results

118 ## *Planococcus citri* short- and long-read transcriptome resource

119 To complement the available *P. citri* genomic sequence data (*Pcitri*.v1 genome assembly) we
120 generated short and long transcriptome reads by performing RNA-Seq and Iso-Seq. Short
121 Illumina reads from virgin and mated *P. citri* females generated in this study were *de novo*
122 assembled on its own (Figure 1a), and together with other available *P. citri* short-read sequence
123 data, including reads from *P. citri* males (Figure 1b). Virgin and mated female-specific
124 transcriptomic data was generated to obtain information on possible pheromone
125 synthesis-related changes in gene expression. Both assemblies were combined, resulting in
126 389,962 unique sequences (Figure 1c). After the merge with the long-read transcriptome
127 (Figure 1d), the full consolidated set contained 440,881 unique transcript sequences
128 (Figure 1e), with an average length of 1,356 nucleotides and N50 of 2,908 nucleotides
129 (Table S1). On average, around 80 % of short reads from each sample mapped back to *de novo*
130 transcriptome assembly (Table S2). To retain as much variability as possible, we did not do
131 any further assembly thinning and the presented dataset is, therefore, redundant.
132 Approximately 53 % of the transcripts mapped to the *Pcitri*.v1 genome, with a third of them
133 mapping to the annotated *Pcitri*.v1 gene models. Completeness assessment with BUSCO found
134 98.7 % complete BUSCOs for the consolidated transcriptome, the majority of which were
135 duplicated (Figure S1). Transcript IDs with their InterPro annotations, information on mapping
136 to *Pcitri*.v1 genome and virgin versus mated female differential expression values are available
137 in Table S3a, as well as on FAIRDOMHub along with a fasta file with all sequences from the
138 consolidated transcriptome (see Data Availability Statement).

139

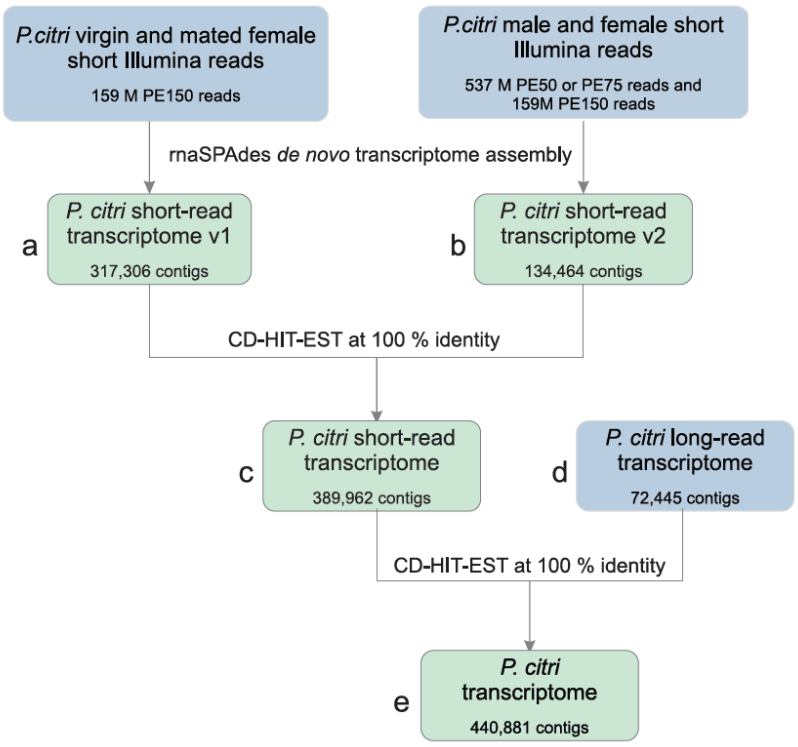

Figure 1: Construction of *P. citri* transcriptome resource. Diagram of consolidation steps taken to combine the short-read (a, b, merged into c) and long-read (d) *P. citri* transcriptome resources into a comprehensive dataset of all assembled and sequenced putative *P. citri* transcripts (e). Raw sequencing data are in blue boxes.

Due to the elusive nature of terpenoid biosynthesis in insects, it has been hypothesised that the origin of some terpenoids (or their precursors), might not be the insects themselves, but rather endosymbiotic bacteria or even plant food (Beran and Petschenka, 2022). For a comprehensive analysis of the biosynthetic capacity for terpenoids found in *P. citri*, our transcriptomic resource, therefore, contains transcripts of all taxonomic origins, as we elected not to exclude non-insect reads. Taxonomic analyses revealed that out of 207,753 protein sequences determined from the consolidated transcriptome, 26 % were unclassified, with the majority of classified sequences assigned to insects (Figure S2). 2,641 sequences were assigned to bacteria, 1,010 of which correlated best with sequences from *Candidatus* Moranella endobia, a secondary endosymbiont of *P. citri*. The acquired data also includes 2,364 viral, 1,629 plant, and 4,299 fungal sequences. Of the plant sequences, most were assigned to potato (*Solanum tuberosum*), probably originating from the *P. citri* food source. Of the fungal sequences, most were assigned to *Wallemia mellicola*, a cosmopolitan fungus with air-disseminated spores (Jančič et al., 2015).

160 **Identification of putative IDS coding sequences**

161 To find sequences with homology to *cis*- and *trans*-IDS enzymes, we searched *Pcitri*.v1
162 genome assembly as well as the *P. citri* consolidated transcriptome (Figure 1e) for sequences
163 with assigned InterPro family annotations IPR001441, and IPR036424 (for *cis*-IDS sequences)
164 or IPR000092 (for *trans*-IDS sequences). On the *Pcitri*.v1 genome assembly we also performed
165 MAST motif search. All extracted sequences (Table S4a and b) were collapsed into clusters
166 based on their sequence similarity. This resulted in a final number of 30 putative IDS sequences
167 (Table S4c). Newly acquired transcript sequence information provided alternative primary
168 sequences to matching *Pcitri*.v1 gene models for seven IDS sequences (Figure 2, Figure S3).
169 For five of them (*cis*IDS1, *trans*IDS4, *trans*IDS5, *trans*IDS11, and *trans*IDS12) we were able
170 to confirm the sequences predicted by the transcriptome data with amplification (Figure 2).

171 Eight of the identified IDS sequences most likely originated from other species: *trans*IDS14
172 has only one mismatch to a FPPS from potato, while *trans*IDS15, *cis*IDS7, and *cis*IDS10 have
173 100 % identity to different IDS sequences from the fungus *Wallemia mellicola*, respectively
174 (Table S4c). *Trans*IDS6, *trans*IDS7, *cis*IDS4, and *cis*IDS6 have homology to insect IDS
175 sequences. However, as determined by phylogenetic analysis (Supplementary Figures S4, S5,
176 S6, and S7), these grouped with homologous sequences from Chalcidoidea wasps
177 (Hymenoptera). Additionally, one sequence can be assigned to a *P. citri* symbiont: *cis*IDS8 is
178 identical to a protein annotated as a UPPS (undecaprenyl diphosphate synthase) from
179 *Candidatus* Moranella endobia.

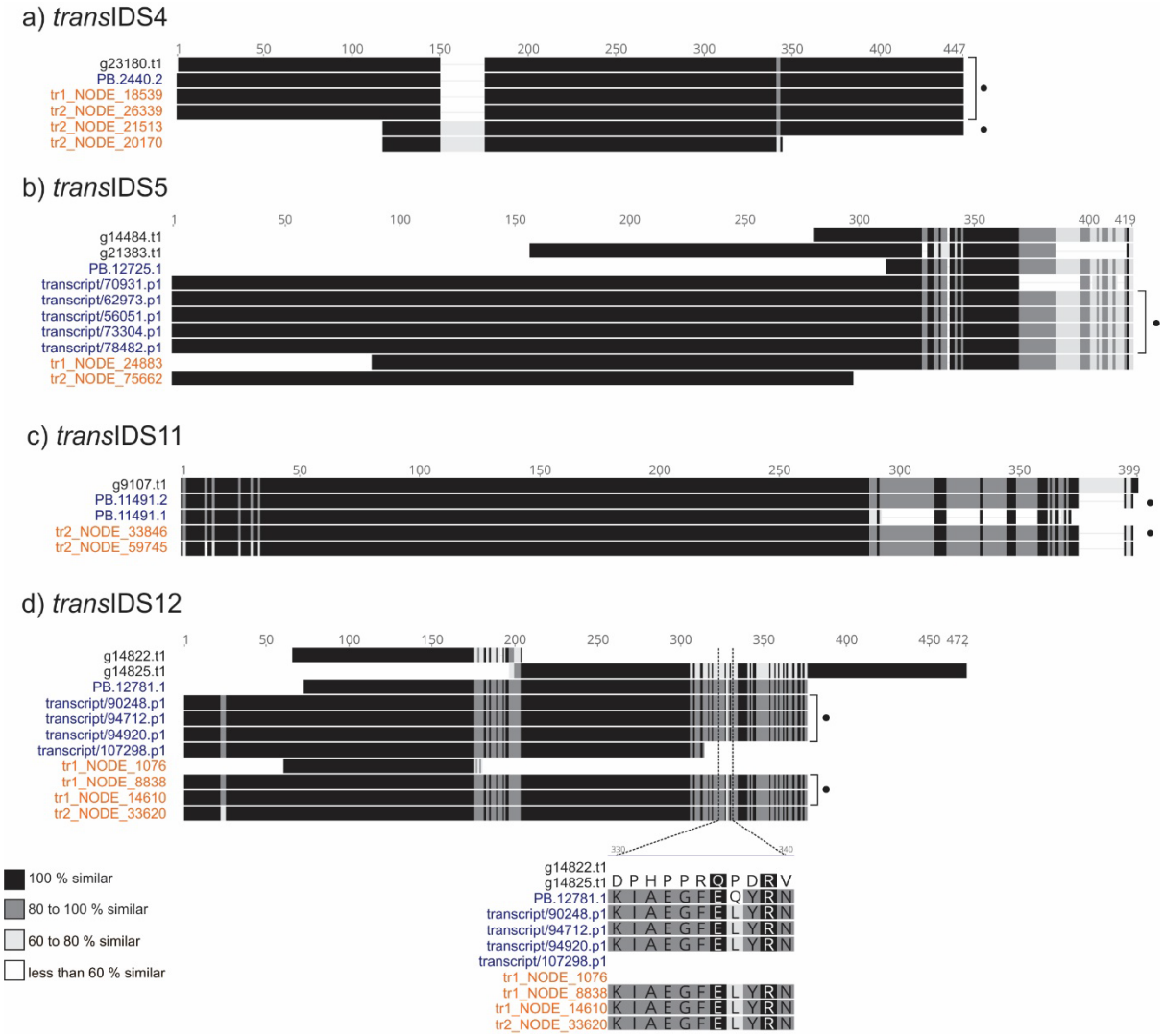

Figure 2: Identification and confirmation of IDS coding sequences predicted by transcriptome data. Multiple sequence alignments for clusters of overlapping transcripts and gene models for candidate sequences *trans*IDS4 (a), *trans*IDS5 (b), *trans*IDS11 (c), and *trans*IDS12 (d). Each cluster includes translated coding sequences obtained from *Pcitri*.v1 gene models (in black text), long-read transcripts (in blue text), and transcripts *de novo* assembled from short reads (in orange text). All sequences are available in Table S4. Sequences confirmed with amplification from cDNA and Sanger sequencing are marked with black dots. Colour-coded sequence similarity is given in the legend in the bottom left. The translated amplicon of *cis*IDS1 has Leu (L327) instead of His at position 327 (H327), the latter predicted in the gene model, while the transcripts all have L327 (not shown in Figure). Two different amplicons were obtained in the case of *trans*IDS4 (a), one identical to the gene model and one with N-terminal truncation and an insertion of 25 amino acids. For *trans*IDS5 (b), the gene model was missing 283 amino acids at the N-terminus. For *trans*IDS11 (c), we amplified a sequence with a frame-shift close to the C-terminus resulting in a 21 amino acids long C-terminus truncation. In the case of *trans*IDS12 (d), two variations of the same full-length sequence were amplified, differing in a single amino acid position: L328 versus Q328. The alignment around the L328 versus Q328 variation is shown below the full length alignments. The structure of the amplified sequence of *trans*IDS12 differed significantly from the annotated gene models. Multiple sequence alignment was done in MEGAX using the MUSCLE algorithm and visualised with Geneious software.

199 ***In silico* analysis reveals diverse putative functions of candidate *P. citri* IDSs**

200 We further analysed the sequence features and phylogeny of identified *P. citri* IDS sequences,
201 which suggested their potential functions (Table 1). Two candidate sequences (*trans*IDS2,
202 *trans*IDS4) are similar to decaprenyl diphosphate synthases (DPPSs), long-chain IDSs that
203 catalyse the formation of the ubiquinone prenyl sidechain, important in aerobic cellular
204 respiration. The functional DPPS enzyme forms a heterodimer of subunits 1 and 2, of which
205 subunit 1 has a higher sequence homology to IDSs, while subunit 2 acts as a regulatory subunit
206 (Zhang and Li, 2013; Song and Li, 2022). T*rans*IDS2 is similar to sequences of DPPS
207 subunit 1, while *trans*IDS4 lacks the D-rich motifs and is similar to sequences of DPPS
208 subunit 2 (Table 1, Supplementary Figure S4, Supplementary Figure S8).

209 *Trans*IDS3 and *trans*IDS5 have high similarity to FPPSs (Table 1, Supplementary Figure S5).
210 Additionally, there are seven sequences (*trans*IDS8, *trans*IDS9, *trans*IDS10, *trans*IDS11,
211 *trans*IDS12, *trans*IDS13, *trans*IDS18) with similarity to FPPS-like sequences from other
212 Coccoidea species (Supplementary Figure S9). All seven of them have a canonical FARM
213 (DDxxD), but not SARM (Table 1). Phylogenetically, the FPPS-like sequences are separated
214 from the mealybug FPPS sequences with low sequence similarity between the two groups
215 (Supplementary Figure S9). Similar diversification pattern of IDS genes within Coccomorpha
216 has been indicated previously (Rebholz et al., 2023). *Trans*IDS3 and *trans*IDS5, as well as
217 *trans*IDS12 and *trans*IDS13 also have high pairwise sequence similarity (78 % and 75 %,
218 respectively), indicating a possibility of more recent duplication events.

219 *Trans*IDS16 is similar to geranylgeranyl pyrophosphate synthases (GGPPS), catalysing the
220 formation of the diterpenoid precursor (Supplementary Figure S10). Sequences of *trans*IDS1,
221 *trans*IDS19, and *trans*IDS20 do not have canonical D-rich motifs. Additionally, *trans*IDS1 is
222 similar to E3 ubiquitin protein ligases, while we could not find any similarity matches for
223 *trans*IDS19 and *trans*IDS20. *Trans*IDS17 is similar to otubain-like ubiquitin thioesterases
224 (Supplementary Figure S11) and includes a DDxxD motif aligning with FARM, as do all other
225 otubain-like orthologs identified from mealybugs and included in the phylogenetic analysis
226 (Supplementary Figure S11).

227 Out of the *cis*-IDS candidates, *cis*IDS1, *cis*IDS2, *cis*IDS3, and *cis*IDS5 are homologous to the
228 catalytic subunit of heterodimeric dehydrodolichyl diphosphate synthase (DHPPS)
229 (Supplementary Figure S6, Supplementary Figure S12), involved in the biosynthesis of
230 dolichol phosphate, the lipid carrier of glycosyl units used for N-glycosylation in eukaryotes.

231 Candidates *cis*IDS2, *cis*IDS3, and *cis*IDS5 have some closer orthologs within the Coccoidea,
232 however they are very distant to other DHPPS sequences and even *cis*IDS1 (Supplementary
233 Figure S12). On the other hand, *cis*IDS9 has highest similarity to sequences of the DHPPS
234 regulatory subunit (Supplementary Figure S7).

235 **Table 1: *P. citri cis*- and *trans*-IDS candidates**. For each candidate sequence, its putative function,
236 differential expression values (logFC), contrasting samples from virgin to mated females, and results of
237 *in vitro* regular (R) and irregular (I) C10 or C15 coupling activity tests of the proteins expressed in *E.*
238 *coli* are given. For *trans*-IDS candidates, amino acids positioned at the first and second D-rich motif
239 (FARM, SARM) according to the multiple sequence alignment are given as well. LogFC (log2-fold
240 changes) values were obtained by averaging logFC values for all sequences in the sequence cluster of
241 each candidate (Table S4c) and are coloured in red and green (for up- and downregulated expression,
242 respectively) if the differential expression was statistically significant (pAdj > 0.05) for at least one of
243 the sequences in each cluster. "/" – activity test not performed. Asterisk denotes the sequence originating
244 from *Candidatus* Moranella endobia.

245

Candidate	FARM	SARM	Putative function	logFC	R	I
*trans*IDS2	DDVID	DDLLD	DPPS subunit 1	0.18	**Y**	N
*trans*IDS3	DDIID	DDYLD	FPPS	1.20	**Y**	N
*trans*IDS4	/	/	DPPS subunit 2	0.32	N	N
*trans*IDS5	DDIMD	DDYLD	FPPS	0.22	**Y**	**Y**
*trans*IDS8	DDILD	NDYND	FPPS-like	-1.26	/	/
*trans*IDS9	DDIFD	NDYND	FPPS-like	3.11	/	/
*trans*IDS10	DDAID	KDQND	FPPS-like	0.93	N	N
*trans*IDS11	DDALD	NDYYD	FPPS-like	1.61	**Y**	N
*trans*IDS12	DDIAD	KLYND	FPPS-like	5.49	N	N
*trans*IDS13	DDVVD	KIFND	FPPS-like	-1.22	/	/
*trans*IDS16	DDIQD	DDYCN	GGPPS	0.08	N	N
*trans*IDS17	DDFHD	/	ubiquitin thioesterase	-0.13	**Y**	N
*trans*IDS18	DDALD	NDYYD	FPPS-like	-0.38	/	/
*cis*IDS1	/	/	DHPPS	0.07	N	N
*cis*IDS2	/	/	DHPPS-like	1.90	/	/
*cis*IDS3	/	/	DHPPS-like	0.09	/	/
*cis*IDS5	/	/	DHPPS-like	0.04	/	/
*cis*IDS8*	/	/	UPPS	NA	N	N
*cis*IDS9	/	/	DHPPS regulatory subunit	0.15	/	/

246

247 **Regular IDS activity was shown for five and irregular activity for one *P. citri* candidate**

248 With the aim of identifying enzymes involved in the synthesis of the irregular C10 sex
249 pheromone, we tested the activity of *P. citri* IDS candidates for producing short polyprenyl
250 chains. We selected 11 *P. citri* candidate sequences with homology to *trans*- or *cis*-IDSs for
251 which the full sequence was confirmed with long-read transcriptome sequencing (Table 1,
252 Table S4c). Two of these sequences (*trans*IDS3 and *trans*IDS10) were synthesised according
253 to long-read sequencing data, and nine sequences (*trans*IDS2, *trans*IDS4, *trans*IDS5,
254 *trans*IDS11, *trans*IDS12, *trans*IDS16, *trans*IDS17, *cis*IDS1, and *cis*IDS8) were amplified from
255 *P. citri* cDNA (Table S4c). All sequences were expressed in *E. coli* and purified recombinant
256 proteins were tested *in vitro* to identify regular and irregular IDS catalytic activity forming C10
257 or C15 backbones.

258 Five *trans*-IDS-like proteins (*trans*IDS2, *trans*IDS3, *trans*IDS5, *trans*IDS11, *trans*IDS17)
259 displayed regular activity (Table 1, Figure 3a). The products of *trans*IDS2, *trans*IDS5,
260 *trans*IDS11, and *trans*IDS17 were identified as geranyl and farnesyl diphosphates based on
261 chromatographic behaviour and mass spectra of diphosphates and their dephosphorylated
262 alcohol derivatives (Figure 3b, Figures S13-S19). Candidate *trans*IDS3 had an unusual pattern
263 of regular products. It formed two isomers of C10 prenyl diphosphate, geranyl and *iso*-geranyl
264 diphosphates, that were further elongated to two C15 isomers: farnesyl and *iso*-farnesyl
265 diphosphate (Figure 3b). Specific activities of *trans*IDS5 were significantly higher than those
266 of the other enzymes, with *trans*IDS3 also exhibiting a relatively high activity for production
267 of C10 prenyl diphosphates (Figure 3a, Table S5). Both *trans*IDS3 and *trans*IDS5 sequences
268 were expressed in *P. citri* females with *trans*IDS3 also differentially expressed between virgin
269 and mated females with higher expression in virgin females (Table 1, Supplementary Table
270 S4c).

271 Moreover, candidate *trans*IDS5 also had a low level of irregular IDS activity, if supplied with
272 DMAPP only (Figure 3b). The enzyme produced two compounds, which were identified as
273 lavandulyl (LPP) and maconelliyl diphosphate (MPP), based on GC-MS spectra of their
274 dephosphorylated derivatives (Figure S20).

275

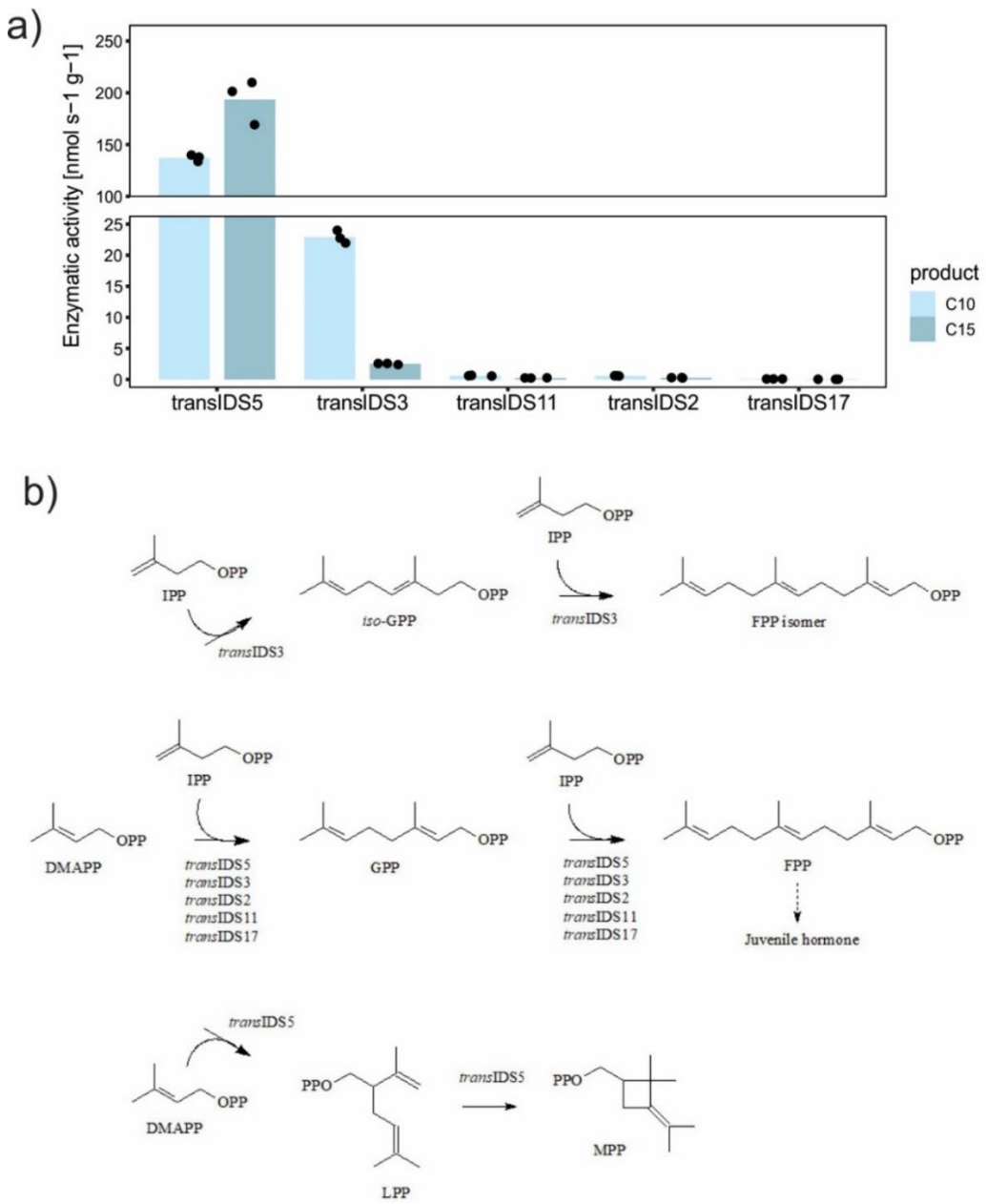

Figure 3: *In vitro* activity of *P. citri* IDS sequences expressed in *E. coli*. Measured enzymatic activity for production of C10 and C15 prenyl diphosphates (a) and schematic representation of confirmed coupling reactions performed by *P. citri trans*-IDS candidates (b). *Trans*IDS5, *trans*ID11, *trans*IDS2, and *trans*IDS17 produced C10 geranyl diphosphate (GPP) and C15 farnesyl pyrophosphate (FPP); *trans*IDS3 produced C10 GPP and *cis*-isogeranyl diphosphate and two C15 FPP isomers. (a) Activity for production of C10 prenyl diphosphates (light blue bars) and activity for production of C15 prenyl diphosphates (dark blue bars). Height of the bars represents the mean of three separate measurements, plotted as block dots. Raw data is available in Table S5. Note the y-axis break due to much higher activity of *trans*IDS5. We did not detect any C10 or C15 products with other tested candidates (Table 1). In (b), dimethylallyl diphosphate (DMAPP) can be joined to its isomer isopentenyl diphosphate (IPP) to form regular terpenes geranyl and iso-geranyl diphosphates (GPP and iso-GPP) and isomers of farnesyl diphosphate (FPP). Alternatively, two DMAPP units can be assembled into irregular lavandulyl diphosphate (LPP) and maconelliyl diphosphate (MPP). FPP is a precursor for juvenile hormone biosynthesis as denoted by the dashed arrow.

291 To assess the activities of candidate proteins *in vivo* in the eukaryotic cell environment, the
292 candidate IDS sequences were transiently expressed in the plant heterologous expression model
293 *Nicotiana benthamiana*. HPLC-MS analysis of leaf extracts did not reveal irregular terpene
294 diphosphates. Similarly, no production of volatile organic compounds derived from irregular
295 monoterpenoids was detected by GC-MS in agroinfiltrated *N. benthamiana* leaves. Regular
296 activity of candidate proteins could not be assessed in this system due to the presence of native
297 plant regular monoterpenes.

298 **Mutagenesis of *trans*IDS5 active site revealed functional importance of active site residues**
299 **for its enzymatic activity**

300 We further set to identify the amino acid residues important for the irregular coupling activity
301 of *trans*IDS5 by performing mutagenesis. It was suggested that replacement of the first
302 aspartate (D) in SARM with asparagine (N) played an important role in the evolution of
303 irregular coupling activity (Liu et al., 2012). However, in *trans*IDS5, analogous mutation
304 (D308N) caused a complete loss of irregular activity. The same was observed after the
305 exchange of two other aspartate residues in SARM (D309N and D312N), and the first aspartate
306 in FARM (D166N).

307 Moreover, the regular activity of aspartate mutants was affected to different degrees
308 (Figure 4a). The regular activity of D166N mutant was significantly diminished, although the
309 ratio between C10 GPP and C15 FPP was similar to the wild-type (wt) *trans*IDS5 (Figure 4a).
310 D308N and D312N substitutions had impact on the regular product length. Both mutants
311 performed the initial coupling of IPP and DMAPP to C10 chains at levels comparable to the
312 wt enzyme, but the addition of the second IPP to form the C15 chain was hampered compared
313 to the wt (Figure 4a). The regular activity was completely abolished after the D309N exchange
314 (Figure 4a).

315

316

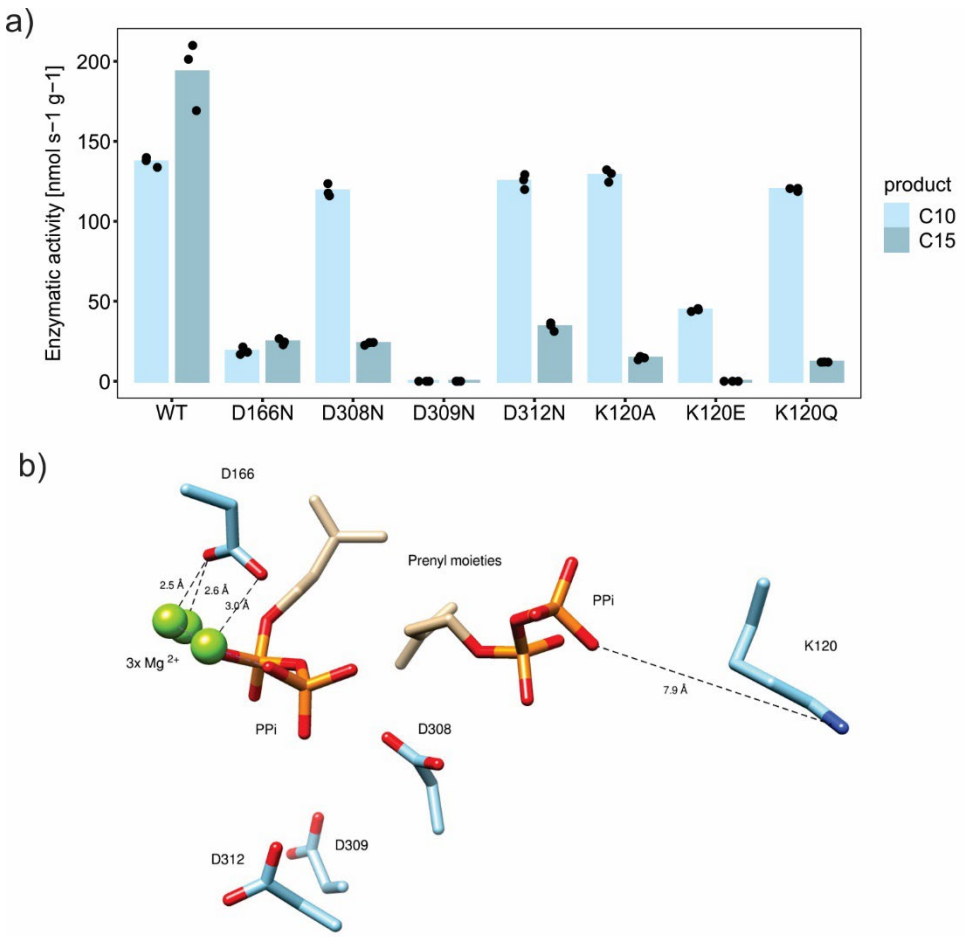

Figure 4: Mutagenesis of *trans*IDS5 active site residues. (a) Activity of *trans*IDS5 and its mutants expressed in *E. coli.* Activity for production of C10 geranyl diphosphate (GPP, light blue bars) and activity for production of C15 farnesyl diphosphate (FPP, dark blue bars) are plotted. Height of the bars represents the mean of three separate measurements, plotted as black dots. Raw data is available in Table S5. (b) Mutated aminoacids and their position in the active site, based on a computational model of active centre of *trans*IDS5 with two bound substrate molecules showing the aspartate of the first (D166) and the second (D308, D309, D312) aspartate-rich motifs and lysine 120 (K120). Oxygen atoms in aspartate side chains and in pyrophosphate moieties (PPi) of the substrates are shown in red; the nitrogen atom of lysine side chain is highlighted in blue.

To find additional residues important for the catalytic activity of *trans*IDS5, we build a computational model of its 3D structure (Figure 4b). We identified a positively charged lysine (K) residue in position 120 located at a comparatively short distance from the negatively charged diphosphate moiety of the substrate (Figure 4b). We exchanged this residue with alanine (K120A), polar uncharged glutamine (K120Q), and negatively charged glutamate (K120E). All three substitutions resulted in the loss of irregular coupling activity, as well as in the diminished elongation of GPP (C10) to FPP (C15). As expected, the K120E exchange, introducing negative charge in the active site, had the most inhibiting effect, precluding FPP formation entirely (Figure 4a).

336 **Heteromeric associations did not change IDS activity**

337 In most cases, *trans*- and *cis*-IDSs are active in the form of homodimers (Gao et al., 2012).
338 However, some IDSs form heterodimers with non-catalytic subunits, which modulate or
339 stimulate their activity (Burke et al., 1999; Tholl et al., 2004; Qu et al., 2015; Yin et al., 2017;
340 Zhou et al., 2017; Emi et al., 2019; Dong et al., 2023). To determine if irregular coupling might
341 depend on the presence of other subunits, we tested *P. citri* IDS candidates in combinations
342 with potential interactors (Table S6). In the *P. citri* genome, we identified homologs of the
343 DPPS and DHPPS regulatory subunits (*trans*IDS4 and *cis*IDS9, respectively) and both were
344 combined with all IDS-like candidates in irregular activity assays (Table S6). In addition, we
345 examined two *P. citri* non-IDS-like proteins for their IDS-modifying activities: the protein
346 product of g23689.t1, annotated as a pheromone-binding protein (PBP) and highly expressed
347 in virgin females (Table S3b), and g36424.t1, identified as a *P. citri* homolog of isopentenyl
348 diphosphate isomerase (IDI), an enzyme catalysing the reversible isomerisation of IPP to
349 DMAPP (Table S6). Addition of *P. citri*-IDI to the reaction with IDS candidates supported
350 formation of regular products, if either IPP or DMAPP only were added to the reaction.
351 However, none of the tested putative regulatory subunits altered or spurred irregular coupling
352 activity of IDS candidates.

353 We also tested the hypothesis that heterodimeric complexes consisting of two IDS subunits are
354 required for irregular coupling. Activity assays with DMAPP as the only substrate were carried
355 out with pairwise combinations of *P. citri* IDS candidate proteins with each other as well as
356 with IDS proteins forming irregular prenyl diphosphates (Table S6). The latter allows for
357 detection of terpene synthase (TPS) activities of *P. citri* IDS candidates, converting the prenyl
358 diphosphate substrates into corresponding prenyl alcohols. Combined assays did not reveal any
359 changes to the enzymatic activities of IDS candidates, which also did not demonstrate TPS
360 activity on the tested irregular prenyl diphosphates.

361

362 ## Discussion

363 Sustainable insect pest management, supported by state-of-the-art knowledge and technology,
364 is an integral part of sustainable food production in a changing climate, in which the
365 geographical ranges of many insect species are expected to widen (Schneider et al., 2022).
366 However, the use of biotechnology and genetic solutions depend on detailed information about
367 gene function. In this work, we present the results of a comprehensive search for candidate
368 genes coding for isopentenyl diphosphate synthase (IDS) activity from the citrus mealybug, *P.*
369 *citri*. The identification of these genes has the potential to enable either biological production
370 of *P. citri* sex pheromones or gene silencing approaches, both of which provide novel options
371 for pest management (reviewed in Mateos Fernández et al., 2022).

372 Functional genomics of many insect species is hindered by the lack of well-annotated, high-
373 quality genome sequences (Li et al., 2019; Hotaling et al., 2021). To improve the available
374 genetic resources for *P. citri,* we complemented the existing genome data with short- and
375 long-read transcriptome data (Figure 1). The transcriptomic dataset enabled us to predict and
376 amplify alternative transcript sequences (Figure 2, Figure S3). In addition, we found a novel
377 coding sequence (*trans*IDS5). Using all resources, we were able to identify 18 candidate *P. citri*
378 IDS sequences. Among these were putative DHPPS, FPPS, FPPS-like, GGPPS, DPPS and
379 DPPS-like coding sequences, as well as an ubiquitin thioesterase-like sequence with a D-rich
380 motif (Table 1).

381 However, inferring function from sequence analyses alone is inconclusive for IDSs, as even
382 small changes can alter substrate and chain-length specificity (Wallrapp et al., 2013). To test
383 the activity of IDS candidates, we focused on the detection of C10 and C15 coupling products.
384 We, therefore, did not confirm if the predicted DHPPS, GGPPS and DPPS had longer
385 chain-producing activities. *Trans*IDS16, a predicted GGPPS did not produce any C10 or C15
386 products. Similarly, neither of the two tested *cis*-IDS-like proteins produced detectable levels
387 of regular or irregular C10 or C15 backbones, as both were predicted to be long-chain IDSs,
388 namely DHPPS (*cis*IDS1) and UPPS (*cis*IDS8) (Table 1). Interestingly, *trans*IDS2, a predicted
389 long-chain IDS (DPPS catalytic subunit), produced low amounts of FPP and GPP. This might
390 be attributed to functional promiscuity *in vitro*, producing polyprenyl chains of different
391 lengths, as reported for other IDS enzymes (Lackus et al., 2019). Additionally, *trans*IDS17,
392 which was not homologous to IDS sequences but rather to otubain-like ubiquitin thioesterases,
393 was able to produce regular C10 and C15 backbones, albeit at the lowest rate of all candidates.

394 It remains unknown if otubain-like sequences from other mealybug species have IDS activity,
395 and whether this activity has a biological function.

396 The two putative FPPS sequences (*trans*IDS3 and *trans*IDS5) both displayed regular coupling
397 activity resulting in C10 and C15 units. Our enzymatic activity and gene expression results
398 indicate that *trans*IDS5 might be the major source of C15 terpenes in *P. citri*, while the C10
399 could be mainly composed of *trans*IDS5 and possibly also *trans*IDS3 products. The highest
400 specific activity was measured for *trans*IDS5, also producing low amounts of irregular prenyl
401 diphosphates, LPP and MPP (Figure 3b), which have not been reported in *P. citri*. The
402 2-methylbutanoated versions of both compounds have been identified as the sex pheromone of
403 *Maconellicoccus hirsutus*, the pink hibiscus mealybug (Zhang et al., 2004). Esterified
404 lavandulol compounds act as sex pheromones also in some other *Planococcus*, *Pseudococcus*
405 and *Dysmicoccus* species (Zou and Millar, 2015). However, the irregular coupling activity of
406 *trans*IDS5 only takes place in the absence of IPP and might, therefore, be unlikely to occur
407 *in vivo* where the presence of IPP is expected. Although we did not find conclusive evidence
408 for the role *trans*IDS5 in the biosynthesis of the *P. citri* sex pheromone, we propose that it is
409 the main source of regular sesquiterpenes in *P. citri*, possibly including the juvenile hormone
410 (Figure 3b). Using structure-informed mutagenesis, we additionally demonstrated the
411 importance of K120, D166, D308, D309, and D312 for the irregular coupling mechanism
412 observed for *trans*IDS5 *in vitro* (Figure 4a).

413 Terpenoids are known for their extensive evolutionary divergence, with many instances of
414 lineage-specific pathways and, therefore, metabolites. Despite the occurrence of many
415 common and unique terpenoids in animals, terpenoid diversification is better understood in
416 plants and microbes (Beran et al., 2019; Tholl et al., 2023). In plants, terpenoid diversity is
417 mainly determined by the activity of TPSs, which have high evolutionary divergence and
418 functional plasticity (Karunanithi and Zerbe, 2019). However, sequences with homology to
419 plant and microbial TPSs have not been identified in insect genomes. Instead, numerous
420 instances of IDS gene-family expansion have been observed, with frequent gene duplications
421 resulting in functional redundancy or neofunctionalization (Gilg et al., 2009; Beran et al., 2016;
422 Lancaster et al., 2018, 2019; Darragh et al., 2021; Rebholz et al., 2023). Among the *P. citri*
423 IDS sequences, we observed expansions of FPPS-like sequences and DHPPS-like sequences
424 (Figure S9, Figure S12). The former extends the previously reported FPPS-like diversification
425 within the Coccomorpha (Rebholz et al., 2023). Three FPPS-like sequences were successfully
426 cloned and tested (*trans*IDS10, *trans*IDS11, and *trans*IDS12), however, only one candidate

427 (*trans*IDS11) exhibited IDS activity, producing low amounts of GPP and FPP. Interestingly,
428 while we were not able to confirm GPPS or FPPS activity for *trans*IDS12, its expression was
429 observed to be strongly upregulated in virgin compared to mated *P. citri* females, which was
430 also true for *trans*IDS9, *trans*IDS10, and *trans*IDS11 (Table 1). This differential expression
431 suggests a possible role in mating or reproduction and needs further investigation.

432 Despite detecting two irregular prenyl diphosphate products, none of the IDS-like candidates
433 that we tested displayed coupling activity leading to *cis*-planococcyl diphosphate under our
434 experimental conditions. *In vitro* functional characterization of IDS enzymes can be unreliable,
435 as activity can depend on many factors, including the formation of heteromers (Zhang and Li,
436 2013). To that end, we tested for modulating activity of several potential interactors but did not
437 detect any changes in the activity of *P. citri* IDS candidates.

438 The model of irregular terpenoid biosynthesis via a genome-encoded IDS has been challenged.
439 Alternative hypotheses include the possibility that insects sequester plant-derived compounds
440 or utilise products of their endosymbionts' metabolism. Many studies reveal a straightforward
441 strategy of modification and accumulation of plant defence compounds (Beran and Petschenka,
442 2022), but, in some cases, the derivatives of plant secondary metabolites are also used in sexual
443 communication. For example, fragments of plant pyrrolizidine alkaloids and phenylpropanoid
444 metabolites are reported to make danaine butterfly males more attractive to the co-specific
445 females (Nishida et al., 1991). However, the presence of irregular monoterpene structures has
446 not been reported for known *P. citri* host plants which makes the sequestration hypothesis
447 improbable. Like other insects feeding on nutritionally unbalanced plant sap, mealybugs rely
448 on obligate endosymbiotic bacteria as a source of amino acids and vitamins (Husnik et al.,
449 2013). *P. citri* is a part of a nested tripartite endosymbiotic system, where the insect harbours
450 the Betaproteobacteria *Candidatus* Tremblaya princeps, which contains the
451 Gammaproteobacteria *Candidatus* Moranella endobia (McCutcheon and Dohlen, 2011; López-
452 Madrigal et al., 2013). Although we were unable to identify any short-chain IDS-coding
453 sequences within the genomes of *P. citri* endosymbionts, we cloned and tested one IDS
454 sequence, a putative UPPS, from *Candidatus* Moranella endobia (*cis*IDS8). However, this did
455 not produce any regular or irregular short prenyl chains. Moreover, genetic crosses between
456 *P. citri* and *Planococcus minor* - the latter producing a branched lavandulol-type
457 monoterpene - indicate that the enzyme responsible for the formation of the irregular terpene
458 skeleton is encoded in the insect nuclear genome and is present at a single locus (Tabata, 2022).

459 The structure of the sex pheromone is also not inherited maternally (Tabata, 2022), as would
460 be expected if synthesis requires the transfer of egg-transmitted endosymbionts.

461 Another possibility is that the mealybug sex pheromones are biosynthesised via a different
462 catalytic mechanism, executed by a different, and as yet unidentified, class of enzymes. Novel
463 terpene synthases have been described, expanding on the canonical models of terpene
464 biosynthesis (Rudolf and Chang, 2020). A recent example is the elucidation of the iridoid
465 pheromone found in the pea aphid, *Acyrthosiphon pisum*. Biosynthesis of this molecule
466 proceeds through the same intermediates as in plants, however, it was shown to involve a set
467 of enzymes which lack homology to the plant counterparts (Köllner et al., 2022). The insect
468 enzymes were identified by transcriptomic analysis of the female hind legs, the site of
469 pheromone biosynthesis (Köllner et al., 2022). Targeted transcriptome analyses of pheromone
470 glands were also useful for determining the genes related to pheromone metabolism and
471 transport in Lepidoptera species (Vogel et al., 2010; Gu et al., 2013; Nuo et al., 2021; Yao et
472 al., 2021). At present, however, such approaches are not possible in *P. citri* as the site of the
473 pheromone gland remains unknown.

474 The identification of the *P. citri* IDS gene family provides an important foundation for
475 deciphering terpenoid biosynthesis in this species as well as other mealybugs. The candidate
476 FPPS and FPPS-like sequences with confirmed regular and irregular activity might also present
477 a starting point for protein engineering; mutations have been shown to increase or even spur
478 irregular coupling activity in IDSs (Chan et al., 2017; Gerasymenko et al., 2022).
479 Diversification of IDS sequences in Coccoidea is an intriguing genomic enigma for future
480 investigation, especially in relation to their capacity for species-specific irregular monoterpene
481 synthesis, a unique feature in this insect superfamily, which still remains elusive.

482

483 # Methods

484 **Insect rearing and sampling**

485 The stock colony of *P. citri* was established in the insectary facilities at the Universitat
486 Politècnica de València (UPV, Valencia, Spain) using specimens from Servei de Sanitat
487 Vegetal of Generalitat Valenciana (GVA, Valencia, Spain). Mealybugs were reared on organic
488 green lemons, maintained in a rearing chamber, in the dark at 24 ± 2 °C, with 40−60 % relative
489 humidity. Mated females were allowed to oviposit and, when ovisacs were laid, they were
490 gently transferred with an entomological needle to new lemons. To obtain samples of virgin
491 females, some lemons were separated from the main colony and were visually inspected every
492 3−4 days for the presence of male cocoons, which were manually removed with an
493 entomological needle to leave only virgin females. Mated females were sampled on the lemons
494 from the main stock colony after checking for the presence of ovisacs. The production of sex
495 pheromone was confirmed by collection of emitted volatiles from a group of 20−30 virgin
496 females by solid phase microextraction technique (SPME) (Vacas et al., 2017). The identity of
497 the sex pheromone was unequivocally confirmed by full coincidence of its mass spectra and
498 retention time with an analytical standard sample synthesised according to the method
499 described by Kukovinets et al., 2006.

500 **RNA isolation**

501 RNA was isolated from approximately 150 (cca. 100 mg) pooled virgin and mated *P. citri*
502 females separately, using TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, USA)
503 according to the manufacturer's instructions. RNA was isolated from four biological replicates,
504 resulting in eight RNA samples, four from pools of virgin females and four from pools of mated
505 females. RNA quality was assessed using the Quant-iT™ RNA Assay Kit (Thermo Fisher
506 Scientific, Waltham, MA, USA).

507 **Short read Illumina RNA-Seq**

508 Approximately 1 µg of isolated RNA from all eight RNA samples was purified to extract
509 polyadenylated mRNA using biotin beads, fragmented, and primed with random hexamers to
510 produce the first cDNA strand followed by second-strand synthesis to produce double-stranded
511 cDNA. Libraries were constructed at the Earlham Institute, UK, on a Sciclone G3 NGSx
512 workstation (PerkinElmer, Waltham, MA, USA) using the TruSeq RNA protocol v2 (Illumina
513 Part # 15026495 Rev.F) and sequenced on two lanes of a HiSeq 4000 (Illumina, San Diego,

514 CA, USA) generating 150 bp negative strand-specific paired-end reads. Illumina reads were
515 subjected to adapter trimming, read quality filtering, mapping and read summarization
516 (counting only uniquely mapped paired reads) in CLC Genomics Workbench 10.0.1 (Qiagen,
517 Hilden, Germany), with mapping parameters: mismatch cost 2, insertion cost 3, deletion cost
518 3, length fraction 0.9, similarity fraction 0.9. Reads were mapped to the *Pcitri*.v1 genome,
519 downloaded from the MealyBugBase (https://ensembl.mealybug.org/index.html). Differential
520 expression analysis was performed with R packages edgeR and limma (Law et al., 2018),
521 contrasting samples from virgin and mated *P. citri* females.

522 **Long-read Iso-Seq transcriptome**

523 Approximately 2 µg of RNA isolated from pooled virgin *P. citri* females was sent for PacBio
524 SMRT sequencing using the Iso-Seq protocol (National Genomics Infrastructure, Sweden).
525 High and low quality full-length transcript isoforms were mapped to the *Pcitri*.v1 genome
526 using minimap2 (Li, 2018). Based on the genome mapping, the isoforms were collapsed and
527 representative isoforms were filtered to exclude isoforms with 5' artefacts, all using scripts
528 from cDNA_cupcake Github repository (Tseng, 2022).

529 ***De novo* short-read transcriptome assembly**

530 Trimmed paired and orphan Illumina RNA-Seq reads were subjected to *de novo* transcriptome
531 assembly using rnaSPAdes v3.13.0 (Bushmanova et al., 2019). Two assemblies were created,
532 first, using the virgin and mated female *P. citri* paired-end short reads (2x150 nt) obtained in
533 this study, and second, combining them with paired-end reads of seven samples of *P. citri,*
534 kindly provided by Dr. Laura Ross at the University of Edinburgh (2x75 nt and 2x50 nt, see
535 Data Availability Statement). For the first assembly, default rnaSPAdes parameters were used,
536 whereas for the second assembly, adapter trimming and sequencing artefact removal was
537 additionally performed using BBTools' bbduk script (Bushnell, 2022), followed by SPAdes'
538 BayesHammer (Nikolenko et al., 2013) and BBTools' ecco, ecc (6 passes) and tadpole read
539 error correction scripts. Assembly was done using k-mer lengths 29 and 49.

540 **Consolidation of sequence resources, annotation and quality control**

541 The two Illumina short read assemblies (Figure 1a and b) were combined into a single set
542 (Figure 1c) by removing identical sequences with CD-HIT-EST v6.4 (Fu et al., 2012) using
543 100% sequence identity threshold. In a successive run, the combined short read transcriptome
544 and the Iso-Seq transcriptome (combining mapped and unmapped representative isoforms,

545 Figure 1d) were consolidated into a single set (Figure 1e), again using CD-HIT-EST at 100%
546 sequence identity threshold. The final consolidated transcriptome dataset as well as both
547 short- and long-read datasets were quality checked using rnaQUAST (Bushmanova et al.,
548 2016) and their completeness was assessed using BUSCO v5.4.3 (Manni et al., 2021) with
549 insecta_odb10 lineage dataset. The consolidated transcriptome set was mapped to the *Pcitri*.v1
550 genome using STARlong v2.7.5c (Dobin et al., 2013) and mapped transcripts were matched
551 with the gene models and scaffolds using matchAnnot (Skelly, 2015; version 20150611.02).
552 Transcripts were also subjected to ORF prediction and translation using Transdecoder v5.5.0
553 (Haas, 2023), followed by the annotation of the predicted protein sequences with PFAM and
554 InterPro (IPR) IDs using InterProScan v5.56-89.0 (Jones et al., 2014). The longest ORFs were
555 further used for taxonomic classification using mmseq2 taxonomy tool (Mirdita et al., 2021),
556 querying against the non-redundant NCBI protein database (downloaded on 19.07.2022).
557 Results of taxonomic classification were visualised with Pavian (Breitwieser and Salzberg,
558 2020). Illumina short reads were mapped back to both *de novo* assemblies as well as IsoSeq
559 transcriptome using STAR v2.7.5c (Dobin et al., 2013) with default parameters. Short read
560 mapping counts for all three datasets were used for differential transcript expression analysis
561 based on R packages edgeR and limma (Law et al., 2018).

562 **Candidate selection**

563 We queried the IPR annotations of *Pcitri*.v1 gene models and our consolidated transcriptome
564 dataset with IDS-related IPR IDs, namely IPR001441 (Decaprenyl diphosphate synthase-like
565 family), IPR036424 (Decaprenyl diphosphate synthase-like superfamily) and IPR000092
566 (Polyprenyl synthetase family), to identify homologs of *cis*- and *trans*-prenyltransferases.
567 Besides using the IPR search to determine *trans*-IDS homologs, we also conducted a motif
568 search using MAST v5.0.1 in MEME Suite (Bailey et al., 2015). Twenty-one sequences of
569 farnesyl pyrophosphate synthases (FPPS) from different organisms were used as an input for
570 MEME. The identified motifs were used as a query against the *Pcitri*.v1 proteins (Table S7).
571 Additionally, we also searched for *P. citri* isopentenyl diphosphate isomerase (IDI) homologs
572 by querying the *P. citri*.v1 protein sequences with *Nicotiana tabacum* IDI sequence (GenBank
573 accession NP_001313140) using blastp (Altschul et al., 1990).

574 All extracted sequences with homology to *cis*- and *trans*-IDSs longer than 100 amino acids
575 were subjected to CD-HIT with 100% sequence identity threshold to remove duplicates. They
576 were further grouped into clusters, using CD-HIT along with multiple sequence alignments

577 with MUSCLE (Edgar, 2004), and the sequence representing the most probable full-length
578 coding sequence was selected within each cluster.

579 For the phylogenetic analyses of candidate sequences, we first searched for similar sequences
580 using blastp against the non-redundant protein bases or tblastn against the transcriptome
581 shotgun assembly sequences (TSA). Additionally, we searched for homologs from
582 *Planococcus ficus* and *Pseudococcus longispinus* genome assemblies in the MealyBugBase
583 (https://ensembl.mealybug.org/index.html), and from *Maconellicoccus hirsutus* transcriptomic
584 resource (Kohli et al., 2021). Amino acid sequences were aligned using the MUSCLE
585 algorithm (Edgar, 2004) and used to construct maximum-likelihood phylogenetic trees with
586 LG model (Le and Gascuel, 2008) tested with the bootstrap method at 1000 replicates in
587 MEGA-X, version 10.0.5 (Kumar et al., 2018).

588 **Candidate sequence amplification and expression in *E. coli***

589 The candidate sequences were amplified from the cDNA prepared with oligoT primer
590 (T15NNN) from total RNA isolated from virgin *P. citri* females using primers listed in
591 Table S8 and cloned into pMAL-c5X expression vector (New England Biolabs, Ipswich, MA,
592 USA) using *Nco*I/*Bam*HI sites. The proteins were expressed in BL21(DE3)pLysS *E. coli* cells.
593 The expression was induced by 0.5 mM IPTG and carried out at 30 °C overnight. The proteins
594 were purified on the amylose resin (New England Biolabs, Ipswich, MA, USA) and stored
595 at -20 °C in 25 % glycerol. The protein concentration was determined by the method of
596 Bradford (Bradford, 1976).

597 **Enzyme activity assays**

598 Isopentenyl, dimethylallyl, and geranyl diphosphate (IPP, DMAPP, and GPP) were purchased
599 as tri-ammonium salts from Echelon Biosciences (Salt Lake City, UT, USA). The compounds
600 were dissolved in the mixture of methanol:water (7:3) at the concentration of 1 mg/mL and
601 stored at -20 °C. The activity assays were carried out using 10–180 µg of protein in a total
602 volume of 100 µL in 35 mM HEPES buffer, pH 7.4, containing 10 mM $MgCl_2$ and 5 mM β-
603 mercaptoethanol at 30 °C for 30 or 60 min. Regular coupling activity was determined using
604 IPP and DMAPP as substrates, both at 100 µM, while the assays for irregular activity contained
605 200 µM of only DMAPP. The reaction was stopped by adding 100 µL of chloroform, and the
606 proteins were precipitated by vigorous shaking followed by centrifugation (10 min at 16,000
607 g). The water phase (injection volume 5 µL) was used for LC-MS analysis performed on an
608 1260 Infinity HPLC system coupled to a G6120B quadrupole mass spectrometry detector

609 (Agilent Technologies, Santa Clara, CA, USA). The separation was carried out on a Poroshell
610 120 EC-C18 (3.0 x 50 mm, 2.7 μm) column (Agilent Technologies, Santa Clara, CA, USA)
611 using the method described in (Nagel et al., 2012). The detection of C10 and C15 prenyl
612 diphosphates was performed in negative single ion mode; m/z 313 and 381, respectively. The
613 quantity of formed products was calculated based on peak areas using a calibration line built
614 for GPP.

615 For analysis of regular terpene alcohols, the reactions were carried out with 70–330 μg of
616 protein in a total volume of 400 μL in 35 mM HEPES buffer, pH 7.4, containing 10 mM $MgCl_2$,
617 5 mM β-mercaptoethanol, and substrates (1 mM IPP and 1 mM DMAPP) at 28 °C overnight.
618 For dephosphorylation of prenyl diphosphates, the reaction mixture was supplemented with 80
619 μL of 0.5 M glycine-NaOH buffer (pH 10.5) containing 5 mM $ZnSO_4$; 80 units of calf intestinal
620 alkaline phosphatase (Promega, Madison, WI, USA) were added and the mixture was incubated
621 at 37 °C for 1 h. The alcohols were extracted 3 times with 400 μL of methyl tert-butyl ether
622 (MTBE); the organic phase was evaporated to approximately 150 μL under compressed air
623 flow. GC-MS analysis was carried out using DB-5ms column (30 m x 0.250 mm x 0.25 μm;
624 Agilent Technologies, Santa Clara, CA, USA) with the following temperature gradient: 50 °C,
625 from 50 °C to 120 °C at 7 °C per min, 120 °C to 320 °C at 35 °C per min, hold 5 min. Carrier
626 gas (H_2) flow was set at 3 mL/min. Injection volume was 1 μL (PTV splitless injection, initial
627 injector temperature 60 °C, final injector temperature 250 °C, injector heating rate 50 °C/sec).
628

629 For identification of irregular products of *trans*IDS5, 1 mg of protein was incubated with 2 mg
630 of DMAPP tri-ammonium salt in 1.6 mL of 35 mM HEPES buffer, pH 7.4, containing 10 mM
631 $MgCl_2$, and 5 mM β-mercaptoethanol at 28 °C overnight. The products were dephosphorylated
632 and extracted as described above, and separated on ALUGRAM® Xtra SIL G/UV_{254} TLC plates
633 (0.20 mm silica gel layer, Macharey-Nagel, Düren, Germany) using hexane:acetone (4:1)
634 mixture. Individual substances were extracted from silica gel with MTBE and analysed by GC-
635 MS using Cyclosil-B capillary column (30 m x 0.250 mm x 0.25 μm; Agilent Technologies,
636 Santa Clara, CA, USA) with the following temperature gradient: 50 °C for 3 min → from 50
637 °C to 120 °C at 2 °C per min → 120 °C to 230 °C at 10 °C per min. Helium was applied as the
638 carrier gas at a flow rate of 1 mL/min. Injection volume was 1 μL (1:100 split).
639

640 **Candidate expression in *Nicotiana benthamiana***

641 Genetic constructs for candidate expression in plants were assembled with the Golden Braid
642 (GB) cloning system (Sarrion-Perdigones et al., 2011). Transcriptional units in alpha-level GB
643 vectors included IDS coding sequences fused with a C-terminal His-tag under control of the
644 35S CaMV promoter. Each candidate gene was tested in its native form and with the addition
645 of a modified chloroplast transit peptide (cTP) of ribulose bisphosphate carboxylase small
646 subunit from *Nicotiana sylvestris* (GenBank accession number XP_009794476.1). If a different
647 (e.g. mitochondrial) localization signal was predicted, it was removed before adding the cTP.
648 Transient expression in *N. benthamiana* leaves was carried out as described earlier
649 (Gerasymenko et al., 2019). For detection of irregular IDS activity, the plant material (200 mg)
650 was frozen in liquid nitrogen, ground and extracted with 80 % methanol (400 µL) by 30 min
651 incubation in the ice ultrasound bath. Supernatants after two centrifugation steps (10 min at
652 16,000 g) were analysed using 1260 Infinity HPLC system coupled to G6120B quadrupole
653 mass spectrometry detector (Agilent Technologies, Santa Clara, CA, USA). The separation was
654 carried out on Zorbax Extend-C18 (4.6 x 150 mm, 3.5 µm) column (Agilent Technologies,
655 Santa Clara, CA, USA) with mobile phase consisting of 5 mM ammonium bicarbonate in water
656 as solvent A and acetonitrile as solvent B applying the following gradient (% B): starting at 0,
657 0 to 8 within 2 min, holding 8 for 15 min, 8 to 64 within 3 min, 64 to 100 within 2 min, holding
658 100 for 4 min, 100 to 0 within 1 min, re-equilibrating at 0 for 12 min. The detection of C10 and
659 C15 prenyl diphosphates was performed in negative single ion mode; m/z 313 and 381,
660 respectively. For detection of volatile compounds, samples from agroinfiltrated leaves were
661 collected 5 dpi, snap-frozen in liquid nitrogen and analysed by GC-MS as described previously
662 (Mateos-Fernández et al., 2021).

663

664 **Computational model and site-directed mutagenesis of *trans*IDS5**

665 The 3-D model of *trans*IDS5 was built on the SWISS-MODEL homology-modelling server
666 (Waterhouse et al., 2018) using the crystal structure of *Saccharomyces cerevisiae*
667 geranylgeranyl pyrophosphate synthase in complex with magnesium and IPP (PDB DOI:
668 10.2210/pdb2E8U/pdb) as a template. For structure visualisation and analysis, UCSF Chimera
669 v.1.14 was applied (Pettersen et al., 2004). Mutations were introduced using the Q5-SDM kit
670 (New England Biolabs) with primers listed in Table S8.

671

Acknowledgements and Funding

All authors gratefully acknowledge the European Research Area Cofund Action 'ERACoBioTech' for the support of the project SUSPHIRE (Sustainable Production of Pheromones for Insect Pest Control in Agriculture), which received funding from the Horizon 2020 research and innovation program under grant agreement No. 722361. M.J., M.P., K.G., and Š.B. acknowledge the financial support from the Slovenian Ministry of Education, Science and Sport, as well Slovenian Research Agency (grant No. P4-0165). I.M.G., E.H., and H.W. acknowledge the support by German Federal Ministry of Education and Research (BMBF), grant number 031B060. K.K. and N.P. acknowledge the support of the UK Biotechnology and Biological Sciences Research Council (BBSRC) Core Strategic Program Grant to the Earlham Institute (Genomes to Food Security; BB/CSP1720/1) and grant BB/R021554/1. S.G. and D.O. acknowledge grant PLEC2021-008020 (PHEROPLUS) by the Spanish Ministry of Science and Innovation, the Next Generation EU initiative and the Spanish Agencia Estatal de Investigación (AEI). S.G. acknowledges a postdoctoral grant (CIAPOS/2021/316) from the Generalitat Valenciana and the Fondo Social Europeo. Illumina sequencing was delivered via the BBSRC National Capability in Genomics and Single Cell Analysis (BBS/E/T/000PR9816) at the Earlham Institute by members of the Genomics Pipelines Group. I.M.G. is grateful to Prof. Dr. Michael Heethoff from the Technical University of Darmstadt and to Prof. Dr. Richard Dehn, Laura Kleditzsch, and Miriam Peters from the University of Applied Sciences in Darmstadt for the possibility to carry out GC-MS measurements. We would also like to thank Dr Laura Ross and her group from the University of Edinburgh for kindly providing their *P. citri* Illumina datasets as well as Dr Rosario Gil Garcia from the University of Valencia for providing insights into the *P. citri* endosymbionts and their genomes.

Data and code availabilitys

Raw Illumina reads from *P. citri* virgin and mated females, along with results of the differential expression analysis (as given in Table S3b as well), were deposited at GEO under accession GSE179660. Raw Iso-Seq reads were deposited at SRA under accession SRR15093694. *P. citri* short reads provided by the University of Edinburgh and used in the second *de novo* transcriptome assembly are deposited at SRA under accession numbers SRR11260462, SRR11260463, SRR11260468, SRR11260469, SRR11260470, and SRR11260471. *De novo* transcriptome assembly FASTA file is available at FAIRDOMHub

704 (https://fairdomhub.org/data_files/6316/content_blobs/17124/download), together with an
705 annotation file (https://fairdomhub.org/data_files/6317/content_blobs/17125/download), also
706 included in the Supplemental Files (Table S3a).

707

708 ## Supplemental Files

709 **Table S1:** Quality assessment of short- and long-read transcriptome data.

710 **Table S2:** Percent of short reads mapping to transcriptome sequences.

711 **Table S3:** Differential expression of *P. citri* genes between virgin and mated females.

712 **Table S4:** Selected IDS sequences.

713 **Table S5:** IDS activity measurements of *P. citri* candidates.

714 **Table S6:** Combined activity assays performed for detection of irregular IDS and terpene
715 synthase activities.

716 **Table S7:** Input and output of the MEME motif search for *P. citri* sequences containing *trans*-
717 IDS motifs.

718 **Table S8:** Primers used for amplification of IDS sequences from *P. citri* cDNA and for
719 introducing mutations into *trans*IDS5 sequence.

720 **Figure S1:** BUSCO assessment of transcriptome completeness.

721 **Figure S2:** Taxonomic classification of the *P. citri* transcriptome dataset.

722 **Figure S3:** Alignments of sequence resources for *trans*IDSl0 and *trans*IDS13.

723 **Figure S4:** Phylogenetic tree of putative DPPS subunit 1 sequences from selected species with
724 *trans*IDS2 and *trans*IDS7.

725 **Figure S5:** Phylogenetic tree of putative FPPS sequences from selected species with
726 *trans*IDS3, *trans*IDS5, and *trans*IDS6.

727 **Figure S6:** Phylogenetic tree of putative DHPPS catalytic subunit sequences from selected
728 species with *cis*IDS1, and *cis*IDS4.

729 **Figure S7:** Phylogenetic tree of putative DHPPS regulatory subunit sequences from selected
730 species with *cis*IDS6, and *cis*IDS9.

731 **Figure S8:** Phylogenetic tree of putative DHPPS regulatory subunit sequences from selected
732 species with *cis*IDS6, and *cis*IDS9.

733 **Figure S9:** Phylogenetic tree of putative FPPS and FPPS-like sequences from Coccoidea
734 species.

735 **Figure S10:** Phylogenetic tree of putative GGPPS sequences from selected species with
736 *trans*IDS16.

737 **Figure S11:** Phylogenetic tree of putative otubain-like sequences from selected species with
738 *trans*IDS17.

739 **Figure S12:** Phylogenetic tree of putative DHPPS and DHPPS-like sequences from selected
740 Coccoidea species.

741 **Figure S13:** Identification of regular monoterpenes synthesised by *trans*-IDS enzymes from
742 *P. citri*.

743 **Figure S14:** EI mass spectra of geraniol generated by dephosphorylation of standard GPP (1A)
744 and products of *trans*IDS5 (1B) and *trans*IDS3 (1C).

745 **Figure S15:** EI mass spectra of geraniol generated by dephosphorylation of standard GPP (1A)
746 and products of *trans*IDS5 (1B) and *trans*IDS3 (1C).

747 **Figure S16:** Identification of regular sesquiterpenes synthesised by *trans*-IDS enzymes from
748 *P. citri*.

749 **Figure S17:** EI mass spectra of *trans*-farnesol generated by dephosphorylation of products of
750 FPPS from *Tanacetum cinerariifolium* known to produce *trans*-farnesyl diphosphate (4A) and
751 *trans*IDS5 (4B).

752 **Figure S18:** EI mass spectra of dephosphorylated products of *trans*IDS3, *trans*-farnesol (4C)
753 and farnesol isomer (5).

754 **Figure S19:** Identification of regular monoterpene diphosphates synthesised by *trans*-IDS
755 enzymes from *P. citri*.

756 **Figure S20:** Identification of irregular monoterpenes synthesised by *trans*IDS5.

757 **Bibliography**

758 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic Local
759 Alignment Search Tool. *J. Mol. Biol.* 215, 403–410. doi:10.1016/S0022-2836(05)80360-
760 2.

761 Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME Suite. *Nucleic*
762 *Acids Res.* 43, W39–W49. doi:10.1093/nar/gkv416.

763 Beran, F., Köllner, T. G., Gershenzon, J., and Tholl, D. (2019). Chemical convergence
764 between plants and insects: biosynthetic origins and functions of common secondary
765 metabolites. *New Phytol.* 223, 52–67. doi:10.1111/nph.15718.

766 Beran, F., and Petschenka, G. (2022). Sequestration of Plant Defense Compounds by Insects:
767 From Mechanisms to Insect-Plant Coevolution. *Annu. Rev. Entomol.* 67, 163–180.
768 doi:10.1146/annurev-ento-062821-062319.

769 Beran, F., Rahfeld, P., Luck, K., Nagel, R., Vogel, H., Wielsch, N., et al. (2016). Novel
770 family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea
771 beetle. *Proc. Natl. Acad. Sci. U. S. A.* 113, 2922–2927. doi:10.1073/pnas.1523468113.

772 Bierl-Leonhardt, B. A., Moreno, D. S., Schwarz, M., Fargerlund, J. A., and Plimmer, J. R.
773 (1981). Isolation, identification and synthesis of the sex pheromone of the citrus
774 mealybug, Planococcus citri (risso). *Tetrahedron Lett.* 22, 389–392. doi:10.1016/0040-
775 4039(81)80107-4.

776 Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram
777 quantitites of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72,
778 248–254. doi:10.1016/0003-2697(76)90527-3.

779 Breitwieser, F. P., and Salzberg, S. L. (2020). Pavian: interactive analysis of metagenomics
780 data for microbiome studies and pathogen identification. 36, 1303–1304.
781 doi:10.1093/bib/bbx120.Kim.

782 Burke, C. C., Wildung, M. R., and Croteau, R. (1999). Geranyl diphosphate synthase:
783 Cloning, expression, and characterization of this prenyltransferase as a heterodimer.
784 *Proc. Natl. Acad. Sci. U. S. A.* 96, 13062–13067. doi:10.1073/pnas.96.23.13062.

785 Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). RnaSPAdes: A de
786 novo transcriptome assembler and its application to RNA-Seq data. *Gigascience* 8, 1–

13. doi:10.1093/gigascience/giz100.

Bushmanova, E., Antipov, D., Lapidus, A., Suvorov, V., and Prjibelski, A. D. (2016). RnaQUAST: A quality assessment tool for de novo transcriptome assemblies. *Bioinformatics* 32, 2210–2212. doi:10.1093/bioinformatics/btw218.

Bushnell, B. (2022). BBMap. Available at: https://sourceforge.net/projects/bbmap/.

Chan, Y.-T., Ko, T.-P., Yao, S.-H., Chen, Y.-W., Lee, C.-C., and Wang, A. H.-J. (2017). Crystal Structure and Potential Head-to-Middle Condensation Function of a Z,Z-Farnesyl Diphosphate Synthase. *ACS Omega* 2, 930–936. doi:10.1021/acsomega.6b00562.

Daane, K. M., Cooper, M. L., Mercer, N. H., Hogg, B. N., Yokota, G. Y., Haviland, D. R., et al. (2021). Pheromone Deployment Strategies for Mating Disruption of a Vineyard Mealybug. *Hortic. Entomol.* 114, 2439–2451. doi:10.1093/jee/toab198.

Darragh, K., Orteu, A., Black, D., Byers, K. J. R. P., Szczerbowski, D., Warren, I. A., et al. (2021). A novel terpene synthase controls differences in anti-aphrodisiac pheromone production between closely related Heliconius butterflies. *PLoS Biol.* 19, e3001022. doi:10.1371/JOURNAL.PBIO.3001022.

Demissie, Z. A., Erland, L. A. E., Rheault, M. R., and Mahmoud, S. S. (2013). The biosynthetic origin of irregular monoterpenes in lavandula: Isolation and biochemical characterization of a novel cis-prenyl diphosphate synthase gene, lavandulyl diphosphate synthase. *J. Biol. Chem.* 288, 6333–6341. doi:10.1074/jbc.M112.431171.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635.

Dong, C., Zhang, M., Song, S., Wei, F., Qin, L., Fan, P., et al. (2023). A Small Subunit of Geranylgeranyl Diphosphate Synthase Functions as an Active Regulator of Carotenoid Synthesis in Nicotiana tabacum. *Int. J. Mol. Sci.* 24, 992. doi:10.3390/ijms24020992.

Dunkelblum, E., Zada, A., Gross, S., Fraistat, P., and Mendel, Z. (2002). Sex pheromone and analogs of the citrus mealybug , Planococcus citri : synthesis and biological activity. 25, 1–9.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high

817 throughput. *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.

818 Emi, K., Sompiyachoke, K., Okada, M., and Hemmi, H. (2019). A heteromeric cis-
819 prenyltransferase is responsible for the biosynthesis of glycosyl carrier lipids in
820 Methanosarcina mazei. *Biochem. Biophys. Res. Commun.* 520, 291–296.
821 doi:10.1016/j.bbrc.2019.09.143.

822 Franco, J. C., Cocco, A., Lucchi, A., Mendel, Z., Suma, P., Vacas, S., et al. (2022). Scientific
823 and technological developments in mating disruption of scale insects. *Entomol. Gen.* 42,
824 251–273. doi:10.1127/entomologia/2021/1220.

825 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: Accelerated for clustering the
826 next-generation sequencing data. *Bioinformatics* 28, 3150–3152.
827 doi:10.1093/bioinformatics/bts565.

828 Gao, Y., Honzatko, R. B., and Peters, R. J. (2012). Terpenoid synthase structures: A so far
829 incomplete view of complex catalysis. *Nat. Prod. Rep.* 29, 1153–1175.
830 doi:10.1039/c2np20059g.

831 Gerasymenko, I., Sheludko, Y. V., Navarro Fuertes, I., Schmidts, V., Steinel, L., Haumann,
832 E., et al. (2022). Engineering of a Plant Isoprenyl Diphosphate Synthase for
833 Development of Irregular Coupling Activity. *ChemBioChem* 23, e202100465.
834 doi:10.1002/cbic.202100465.

835 Gerasymenko, I., Sheludko, Y., Fräbel, S., Staniek, A., and Warzecha, H. (2019).
836 Combinatorial biosynthesis of small molecules in plants: Engineering strategies and
837 tools. *Methods Enzymol.* 617, 413–442. doi:10.1016/BS.MIE.2018.12.005.

838 Gilg, A. B., Tittiger, C., and Blomquist, G. J. (2009). Unique animal prenyltransferase with
839 monoterpene synthase activity. *Naturwissenschaften* 96, 731–735. doi:10.1007/s00114-
840 009-0521-1.

841 Gu, S.-H., Wu, K.-M., Guo, Y.-Y., Pickett, J. A., Field, L. M., Zhou, J.-J., et al. (2013).
842 Identification of genes expressed in the sex pheromone gland of the black cutworm
843 Agrotis ipsilon with putative roles in sex pheromone biosynthesis and transport. *BMC*
844 *Genomics* 14, 636. doi:10.1186/1471-2164-14-636.

845 Haas, B. (2023). TransDecoder. Available at:
846 https://github.com/TransDecoder/TransDecoder.

847 Hotaling, S., Sproul, J. S., Heckenhauer, J., Powell, A., Larracuente, A. M., Pauls, S. U., et al.
848 (2021). GBE Long Reads Are Revolutionizing 20 Years of Insect Genome Sequencing.
849 *Genome Biol. Evol.* 13, evab138. doi:10.1093/gbe/evab138.

850 Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., et al. (2013). Horizontal
851 Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite Nested
852 Mealybug Symbiosis. *Cell* 153, 1567–1578. doi:10.1016/j.cell.2013.05.040.

853 Jančič, S., Nguyen, H. D. T., Frisvad, J. C., Zalar, P., Schroers, H. J., Seifert, K. A., et al.
854 (2015). A Taxonomic Revision of the Wallemia sebi Species Complex. *PLoS One* 10,
855 e0125933. doi:10.1371/journal.pone.0125933.

856 Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014).
857 InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30, 1236–
858 1240. doi:10.1093/bioinformatics/btu031.

859 Karunanithi, P. S., and Zerbe, P. (2019). Terpene Synthases as Metabolic Gatekeepers in the
860 Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.* 10, 1166.
861 doi:10.3389/fpls.2019.01166.

862 Kobayashi, M., and Kuzuyama, T. (2018). Structural and Mechanistic Insight into Terpene
863 Synthases that Catalyze the Irregular Non-Head-to-Tail Coupling of Prenyl Substrates.
864 *ChemBioChem* 20, 29–33. doi:10.1002/cbic.201800510.

865 Kohli, S., Gulati, P., Narang, A., Maini, J., Shamsudheen, K. V., Pandey, R., et al. (2021).
866 Genome and transcriptome analysis of the mealybug Maconellicoccus hirsutus:
867 Correlation with its unique phenotypes. *Genomics* 113, 2483–2494.
868 doi:10.1016/j.ygeno.2021.05.014.

869 Köllner, T. G., David, A., Luck, K., Beran, F., Kunert, G., Zhou, J.-J., et al. (2022).
870 Biosynthesis of iridoid sex pheromones in aphids. *Proc. Natl. Acad. Sci. U. S. A.* 119,
871 e2211254119. doi:10.1073/pnas.2211254119.

872 Kukovinets, O. S., Zvereva, T. I., Kasradze, V. G., Galin, F. Z., Frolova, L. L., Kuchin, A.
873 V., et al. (2006). Novel synthesis of Planococcus citri pheromone. *Chem. Nat. Compd.*
874 42, 216–218. doi:10.1007/s10600-006-0082-x.

875 Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular
876 evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–

1549. doi:10.1093/molbev/msy096.

Lackus, N. D., Petersen, N. P., Nagel, R., Schmidt, A., Irmisch, S., Gershenzon, J., et al. (2019). Identification and Characterization of trans -Isopentenyl Diphosphate Synthases Involved in. *Molecules* 24, 2408. doi:10.3390/molecules24132408.

Lancaster, J., Khrimian, A., Young, S., Lehner, B., Luck, K., Wallingford, A., et al. (2018). De novo formation of an aggregation pheromone precursor by an isoprenyl diphosphate synthase-related terpene synthase in the harlequin bug. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8634–E8641. doi:10.1073/pnas.1800008115.

Lancaster, J., Lehner, B., Khrimian, A., Muchlinski, A., Luck, K., Köllner, T. G., et al. (2019). An IDS-Type Sesquiterpene Synthase Produces the Pheromone Precursor (Z)-α-Bisabolene in Nezara viridula. *J. Chem. Ecol.* 45, 187–197. doi:10.1007/s10886-018-1019-0.

Law, C. W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G. K., et al. (2018). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* 5, 1408. doi:10.12688/f1000research.9005.3.

Le, S. Q., and Gascuel, O. (2008). An Improved General Amino Acid Replacement Matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi:10.1093/molbev/msn067.

Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., et al. (2019). Insect genomes : progress and challenges. *Insect Mol. Biol.* 28, 739–758. doi:10.1111/imb.12599.

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi:10.1093/bioinformatics/bty191.

Liu, P.-L., Wan, J.-N., Guo, Y.-P., Ge, S., and Rao, G.-Y. (2012). Adaptive evolution of the chrysanthemyl diphosphate synthase gene involved in irregular monoterpene metabolism. *BMC Evol. Biol.* 12, 214. doi:10.1186/1471-2148-12-214.

López-Madrigal, S., Latorre, A., Porcar, M., Moya, A., and Gil, R. (2013). Mealybugs nested endosymbiosis : going into the ‘ matryoshka ’ system in Planococcus citri in depth. *BMC Microbiol.* 13, 74. doi:10.1186/1471-2180-13-74.

Lucchi, A., Suma, P., Ladurner, E., Iodice, A., Savino, F., Ricciardi, R., et al. (2019). Managing the vine mealybug , Planococcus ficus , through pheromone-mediated mating disruption. *Environ. Sci. Pollut. Res.* 26, 10708–10718.

907 Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021). BUSCO: Assessing
908 Genomic Data Quality and Beyond. *Curr. Protoc.* 1, e323. doi:10.1002/cpz1.323.

909 Mateos-Fernández, R., Moreno-Giménez, E., Gianoglio, S., Quijano-Rubio, A., Gavaldá-
910 García, J., Estellés, L., et al. (2021). Production of Volatile Moth Sex Pheromones in
911 Transgenic Nicotiana benthamiana Plants. *BioDesign Res.* 2021, 9891082.
912 doi:10.34133/2021/9891082.

913 Mateos Fernández, R., Petek, M., Gerasymenko, I., Juteršek, M., Baebler, Š., Kallam, K., et
914 al. (2022). Insect pest management in the age of synthetic biology. *Plant Biotechnol. J.*
915 20, 25–36. doi:10.1111/pbi.13685.

916 McCutcheon, J. P., and Dohlen, C. D. Von (2011). An Interdependent Metabolic Patchwork
917 in the Nested Symbiosis of Mealybugs. *Curr. Biol.* 21, 1366–1372.
918 doi:10.1016/j.cub.2011.06.051.

919 Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J., and Levy Karin, E. (2021). Fast and
920 sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics* 37, 3029–3031.
921 doi:10.1093/bioinformatics/btab184.

922 Nagel, R., Gershenzon, J., and Schmidt, A. (2012). Nonradioactive assay for detecting
923 isoprenyl diphosphate synthase activity in crude plant extracts using liquid
924 chromatography coupled with tandem mass spectrometry. *Anal. Biochem.* 422, 33–38.
925 doi:10.1016/j.ab.2011.12.037.

926 Nagel, R., Schmidt, A., and Peters, R. J. (2019). Isoprenyl diphosphate synthases : the chain
927 length determining step in terpene biosynthesis. *Planta* 249, 9–20. doi:10.1007/s00425-
928 018-3052-1.

929 Nikolenko, S. I., Korobeynikov, A. I., and Alekseyev, M. A. (2013). BayesHammer :
930 Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics* 14,
931 S7.

932 Nishida, R., Kim, C.-S., Fukami, H., and Irik, R. (1991). Ideamine N-Oxides: Pyrrolizidine
933 Alkaloids Sequestered by the Danaine Butterfly, Idea leuconoe. *Biosci. Biotechnol.*
934 *Biochem.* 55, 1787–1792. doi:10.1080/00021369.1991.10870842.

935 Nuo, S.-M., Yang, A.-J., Li, G.-C., Xiao, H.-Y., and Liu, N.-Y. (2021). Transcriptome
936 analysis identifies candidate genes in the biosynthetic pathway of sex pheromones from

937 a zygaenid moth , Achelura yunnanensis (Lepidoptera : Zygaenidae). *PeerJ* 9, e12641.
938 doi:10.7717/peerj.12641.

939 Ogawa, T., Emi, K., Koga, K., Yoshimura, T., and Hemmi, H. (2016). A cis-
940 prenyltransferase from Methanosarcina acetivorans catalyzes both head-to-tail and
941 nonhead-to-tail prenyl condensation. *FEBS J.* 283, 2369–2383. doi:10.1111/febs.13749.

942 Ozaki, T., Zhao, P., Shinada, T., Nishiyama, M., and Kuzuyama, T. (2014). Cyclolavandulyl
943 Skeleton Biosynthesis via Both Condensation and Cyclization Catalyzed by an
944 Unprecedented Member of the cis - Isoprenyl Diphosphate Synthase Superfamily. *J. Am.*
945 *Chem. Soc.* 136, 4837–4840.

946 Passaro, L. C., and Webster, F. X. (2004). Synthesis of the Female Sex Pheromone of the
947 citrus mealybug, Planococcus citri. *J. Agric. Food Chem.* 52, 2896–2899.
948 doi:10.1021/jf035301h.

949 Petkevicius, K., Wenning, L., Kildegaard, K. R., Sinkwitz, C., Smedegaard, R., Holkenbrink,
950 C., et al. (2022). Biosynthesis of insect sex pheromone precursors via engineered β-
951 oxidation in yeast. *FEMS Yeast Res.* 22, foac041. doi:10.1093/FEMSYR/FOAC041.

952 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C.,
953 et al. (2004). UCSF Chimera - A visualization system for exploratory research and
954 analysis. *J. Comput. Chem.* 25, 1605–1612. doi:10.1002/jcc.20084.

955 Qu, Y., Chakrabarty, R., Tran, H. T., Kwon, E.-J. G., Kwon, M., Nguyen, T.-D., et al. (2015).
956 A lettuce (Lactuca sativa) homolog of human Nogo-B receptor interacts with cis-
957 prenyltransferase and is necessary for natural rubber biosynthesis. *J. Biol. Chem.* 290,
958 1898–1914. doi:10.1074/jbc.M114.616920.

959 Rebholz, Z., Lancaster, J., Larose, H., Khrimian, A., Luck, K., Sparks, M. E., et al. (2023).
960 Ancient origin and conserved gene function in terpene pheromone and defense evolution
961 of stink bugs and hemipteran insects. *Insect Biochem. Mol. Biol.* 152, 103879.
962 doi:10.1016/j.ibmb.2022.103879.

963 Rivera, S. B., Swedlund, B. D., King, G. J., Bell, R. N., Hussey, C. E., Shattuck-Eidens, D.
964 M., et al. (2001). Chrysanthemyl diphosphate synthase: isolation of the gene and
965 characterization of the recombinant non-head-to-tail monoterpene synthase from
966 Chrysanthemum cinerariaefolium. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4373–4378.

967 doi:10.1073/pnas.071543598.

968 Rudolf, J. D., and Chang, C.-Y. (2020). Terpene synthases in disguise: enzymology,
969 structure, and opportunities of non-canonical terpene synthases. *Nat. Prod. Rep.* 37,
970 425–463. doi:10.1039/c9np00051h.

971 Sarrion-Perdigones, A., Falconi, E. E., Zandalinas, S. I., Juárez, P., Fernández-del-Carmen,
972 A., Granell, A., et al. (2011). GoldenBraid: An iterative cloning system for standardized
973 assembly of reusable genetic modules. *PLoS One* 6, e21622.
974 doi:10.1371/journal.pone.0021622.

975 Schneider, L., Rebetez, M., and Rasmann, S. (2022). The effect of climate change on
976 invasive crop pests across biomes. *Curr. Opin. Insect Sci.* 50, 100895.
977 doi:10.1016/j.cois.2022.100895.

978 Skelly, T. (2015). MatchAnnot. Available at: https://github.com/TomSkelly/MatchAnnot.

979 Song, X., and Li, Z.-X. (2022). Functional characterization of two different decaprenyl
980 diphosphate synthases in the vetch aphid Megoura viciae. *Arch. Insect Biochem.*
981 *Physiol.* 110, e21900. doi:10.1002/arch.21900.

982 Tabata, J. (2022). Genetic Basis Underlying Structural Shift of Monoterpenoid Pheromones
983 in Mealybugs. *J. Chem. Ecol.* 48, 546–553. doi:10.1007/s10886-021-01339-x.

984 Tholl, D., Kish, C. M., Orlova, I., Sherman, D., Gershenzon, J., Pichersky, E., et al. (2004).
985 Formation of monoterpenes in Antirrhinum majus and Clarkia breweri flowers involves
986 heterodimeric geranyl diphosphate synthases. *Plant Cell* 16, 977–992.
987 doi:10.1105/tpc.020156.

988 Tholl, D., Rebholz, Z., Morozov, A. V., and O'Maille, P. E. (2023). Terpene synthases and
989 pathways in animals: enzymology and structural evolution in the biosynthesis of volatile
990 infochemicals. *Nat. Prod. Rep.* 40, 766–793. doi:10.1039/d2np00076h.

991 Tillman, J. A., Seybold, S. J., Jurenka, R. A., and Blomquist, G. J. (1999). Insect pheromones
992 -an overview of biosynthesis and endocrine regulation. *Insect Biochem. Mol. Biol.* 29,
993 481–514. Available at: www.elsevier.com/locate/ibmb [Accessed May 30, 2023].

994 Tseng, E. (2022). cDNA_Cupcake. Available at:
995 https://github.com/Magdoll/cDNA_Cupcake.

Vacas, S., Alfaro, C., Navarro-llopis, V., and Primo, J. (2010). Mating disruption of California red scale, Aonidiella aurantii Maskell (Homoptera : Diaspididae), using biodegradable mesoporous pheromone dispensers. *Pest Manag. Sci.* 66, 745–751. doi:10.1002/ps.1937.

Vacas, S., Navarro, I., Seris, E., Ramos, C., Hernández, E., Navarro-Llopis, V., et al. (2017). Identification of the male-produced aggregation pheromone of the four-spotted coconut weevil, diocalandra frumenti. *J. Agric. Food Chem.* 65, 270–275. doi:10.1021/acs.jafc.6b04829.

Vogel, H., Heidel, A. J., Heckel, D. G., and Groot, A. T. (2010). Transcriptome analysis of the sex pheromone gland of the noctuid moth Heliothis virescens. *BMC Genomics* 11, 29. doi:10.1186/1471-2164-11-29.

Wallrapp, F. H., Pan, J.-J., Ramamoorthy, G., Almonacid, D. E., Hillerich, B. S., Seidel, R., et al. (2013). Prediction of function for the polyprenyl transferase subgroup in the isoprenoid synthase superfamily. *Proc. Natl. Acad. Sci.* 110, E1196–E1202. doi:10.1073/pnas.1300632110.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., et al. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303. doi:10.1093/nar/gky427.

Watson, G. (2023). CABI Compendum, Planococcus citri (citrus mealybug). Available at: https://www.cabidigitallibrary.org/doi/10.1079/cabicompendium.45082.

Wolk, J. L., Goldschmidt, Z., and Dunkelblum, E. (1986). A short stereoselective synthesis of (+)-cis-planococcyl acetate, sex pheromone of the citrus mealybug Planococcus citri (Risso). *Synthesis (Stuttg).* 1986, 347–348. doi:10.1055/S-1986-31613.

Yao, S., Zhou, S., Li, X., Liu, X., Zhao, W., Wei, J., et al. (2021). Transcriptome Analysis of Ostrinia furnacalis Female Pheromone Gland: Esters Biosynthesis and Requirement for Mating Success. *Front. Endocrinol. (Lausanne).* 12, 736906. doi:10.3389/fendo.2021.736906.

Yin, J.-L., Wong, W.-S., Jang, I.-C., and Chua, N.-H. (2017). Co-expression of peppermint geranyl diphosphate synthase small subunit enhances monoterpene production in transgenic tobacco plants. *New Phytol.* 213, 1133–1144. doi:10.1111/nph.14280.

Zhang, A., Amalin, D., Shirali, S., Serrano, M. S., Franqui, R. A., Oliver, J. E., et al. (2004). Sex pheromone of the pink hibiscus mealybug, Maconellicoccus hirsutus, contains an unusual cyclobutanoid monoterpene. *Proc Natl Acad Sci U S A* 101, 9601–9606. doi:10.1073/pnas.0401298101.

Zhang, H., and Li, Z.-X. (2013). In vitro and in vivo characterization of a novel insect decaprenyl diphosphate synthase: A two-major step catalytic mechanism is proposed. *Biochem. Biophys. Res. Commun.* 442, 105–111. doi:10.1016/j.bbrc.2013.11.025.

Zhou, F., Wang, C.-Y., Gutensohn, M., Jiang, L., Zhang, P., Zhang, D., et al. (2017). A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. *Proc. Natl. Acad. Sci. U. S. A.* 114, 6866–6871. doi:10.1073/pnas.1705689114.

Zou, Y., Chinta, S. P., and Millar, J. G. (2013). "Irregular Terpenoids as Mealybug and Scale Pheromones : Chemistry and Applications," in *Pest Management with Natural Products*, eds. J. J. Beck, J. R. Coats, S. O. Duke, and M. E. Koivunen (American Chemical Society), 125–143. doi:10.1021/bk-2013-1141.

Zou, Y., and Millar, J. G. (2015). Chemistry of the pheromones of mealybug and scale insects. *Nat. Prod. Rep.* 32, 1067–1113. doi:10.1039/c4np00143e.