Iridoid monoterpenes, widely distributed in plants and insects, have many ecological functions. While the biosynthesis of iridoids has been extensively studied in plants, little is known about how insects synthesize these natural products. Here, we elucidated the biosynthesis of the iridoids cis-trans-nepetalactol and cis-trans-nepetalactone in the pea aphid *Acrystashipon pisum* (Harris), where they act as sex pheromones. The exclusive production of iridoids in hind legs of sexual female aphids allowed us to identify iridoid genes by searching for genes specifically expressed in this tissue. Biochemical characterization of candidate enzymes revealed that the iridoid pathway in aphids proceeds through the same sequence of intermediates as described for plants. The six identified aphid enzymes are unrelated to their counterparts in plants, conclusively demonstrating an independent evolution of the entire iridoid pathway in plants and insects. In contrast to the plant pathway, at least three of the aphid iridoid enzymes are likely membrane bound. We demonstrated that a lipid environment facilitates the cyclization of a reactive enol intermediate to the iridoid cyclopentanoid-pyran scaffold in vitro, suggesting that membranes are an essential component of the aphid iridoid pathway. Altogether, our discovery of this complex insect metabolic pathway establishes the genetic and biochemical basis for the formation of iridoid sex pheromones in aphids, and this discovery also serves as a foundation for understanding the convergent evolution of complex metabolic pathways between kingdoms.

Iridoids | aphids | pathway | sex pheromone | biosynthesis

## Significance

Plants, animals, and microbes produce a plethora of natural products that are important for defense and communication. Most of these compounds show a phylogenetically restricted occurrence, but, in rare instances, the same natural product is biosynthesized by organisms in two different kingdoms. The monoterpane-derived iridoids, for example, have been found in more than 50 plant families but are also observed in several insect orders. The discovery of the aphid iridoid pathway, one of the longest and most chemically complex insect-derived natural product biosynthetic pathways reported to date, highlights the mechanisms underlying the convergent evolution of metabolic enzymes in insects and plants, including the recruitment of different enzyme classes to catalyze the same chemical processes.
and plagiodial (11). Feeding experiments with isotopically labeled precursors and the discovery of some of the enzymes involved in chrysomelid formation demonstrated that leaf beetles produce these compounds by a series of chemical reactions similar to those that occur in plants (12–15). Although the enzymatic basis for this pathway has not been completely established, the fact that the known enzymes are unrelated to their counterparts in plants suggests independent evolution of the pathway occurred (14).

**A** Iridoid pathway in plants and aphids

\[ \text{IPPP} + \text{IDS/GPPS} \rightarrow \text{GPP} \]

\[ \text{GES} \rightarrow \text{geraniol} \]

\[ \text{G8H} \rightarrow \text{8-hydroxygeraniol} \]

\[ \text{HGO} \rightarrow \text{8-oxogeraniol} \]

\[ \text{ISY} \rightarrow \text{8-oxocitronellol enol} \]

\[ \text{cyclase} \rightarrow \text{nepetalactol} \]

\[ \text{NEPO} \rightarrow \text{nepetalactone} \]

**B** Relative expression of iridoid pathway genes in *A. pisum*

<table>
<thead>
<tr>
<th>gene ID</th>
<th>gene annotation</th>
<th>f-hl</th>
<th>f-fl</th>
<th>af-hl</th>
<th>m-hl</th>
</tr>
</thead>
<tbody>
<tr>
<td>100162815</td>
<td>acetyl-CoA acetyltransferase</td>
<td>1.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>100165154</td>
<td>HMG-CoA synthase</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>100165462</td>
<td>HMG-CoA reductase</td>
<td>1.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>100163305</td>
<td>mevalonate kinase</td>
<td>1.00</td>
<td>0.10</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>100574505</td>
<td>phosphomevalonate kinase</td>
<td>1.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>100163413</td>
<td>phosphomevalonate kinase-like</td>
<td>1.00</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>100158798</td>
<td>diphosphatevalonate decarboxylase</td>
<td>1.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>10016744</td>
<td>isopentenyl diphosphate isomerase</td>
<td>1.00</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>100144905</td>
<td>isoprenyl diphosphate synthase (ApIDS)</td>
<td>1.00</td>
<td>0.05</td>
<td>0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>100158803</td>
<td>dolichyldiphasphatase 1 (ApIDS)</td>
<td>1.00</td>
<td>0.05</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>100162683</td>
<td>inositol polyphosphate 1-phosphatase</td>
<td>1.00</td>
<td>0.05</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td>100165972</td>
<td>P450 (ApG8H)</td>
<td>1.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>100160284</td>
<td>P450 reductase (ApRed)</td>
<td>1.00</td>
<td>0.11</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>100301633</td>
<td>farnesol dehydrogenase (ApG8H)</td>
<td>1.00</td>
<td>0.02</td>
<td>0.10</td>
<td>0.12</td>
</tr>
<tr>
<td>100162094</td>
<td>retinol dehydrogenase</td>
<td>1.00</td>
<td>0.13</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>103310029</td>
<td>polyrenol reductase (ApISY)</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>100169582</td>
<td>GMC oxidase (ApNEPO)</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>100164798</td>
<td>GMC oxidase</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>100168866</td>
<td>flavin reductase</td>
<td>1.00</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Fig. 1. The formation of iridoids in plants and aphids. (A) Labeling studies suggest that the biosynthesis of iridoids in the pea aphid *A. pisum* mimics the biosynthetic pathway in iridoid-producing plants. IPP, isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP, geranyl diphosphate; IDS, isoprenyl diphosphate synthase; GES, geranyl diphosphate synthase; G8H, geraniol 8-hydroxylase; HGO, 8-hydroxygeraniol oxidoreductase; ISY, iridoid synthase; NEPO, nepetalactone oxidase. (B) Relative expression of mevalonate and putative nepetalactone pathway genes in hind legs and front legs of different sexual stages of *A. pisum*. Relative expression data are based on RPKM values obtained by RNA-seq. f-hl, hind legs of sexual females; f-fl, front legs of sexual females; af-hl, hind legs of asexual females; m-hl, hind legs of males.

The pea aphid *Acyrthosiphon pisum*, for example, has been reported to biosynthesize (1R,4aS,7S,7aR)-cis-trans-nepetalactol and (4aS,7S,7aR)-cis-trans-nepetalactone in glandular structures on the hind legs of sexual female aphids, from where they are released to attract male conspecifics (18, 20). Recent studies with isotopically labeled iridoid precursors suggest that the iridoid pathway in aphids follows the reaction sequence described for plants (21). However, the underlying enzymatic machinery of this pathway is completely unknown.

Here, we report the elucidation of the entire iridoid pathway in the pea aphid *A. pisum*. By searching for genes expressed exclusively in hind legs of sexual female aphids, the site of iridoid production, we could rapidly identify all six biosynthetic genes/enzymes responsible for the conversion of IPP and DMAPP to cis-trans-nepetalactone. The discovery of the insect nepetalactone pathway in its entirety now allows a comparison of the chemical solutions that have evolved for nepetalactone
biosynthesis in plants and animals. Although the chemical steps from GPP to nepetalactone are the same in both *Nepeta* and pea aphids, the enzymes of these pathways have clearly evolved independently.

**Results**

Transcriptome-Enabled Discovery of Iridoid Genes in *A. pism*. Iridoids are produced exclusively in the hind legs of sexual aphid females (20), which allowed us to search for iridoid genes by comparing transcriptomic data from legs of sexual females, asexual females, and males. Since most aphid species have a complex life cycle with multiple asexual generations over the summer and only one sexual female/male generation in fall (22), we subjected a colony of pea aphids to day length/temperature conditions that mimic the fall season to generate sexual females. After verifying that *cis*-trans-nepetalacol and *cis*-trans-nepetalactone were in fact produced in the aphids (*SI Appendix*, Fig. S1), we then collected hind and front legs of sexual females, hind legs of asexual females, and hind legs of males. RNA was extracted and subjected to Illumina sequencing. Out of the 20,918 gene models of the *A. pism* v3 genome, only 96 appeared to be specifically expressed in hind legs of sexual females (*SI Appendix*, Tables S1 and S2). Notably, among these transcripts were 8 genes encoding the entire mevalonate pathway starting from acetyl-CoA to IPP and DMAPP (Fig. 1B). Previ-ously reported isotopic labeling studies indicated that the iridoid pathways in pea aphids and plants share at least some of the same biosynthetic intermediates (21). We therefore assumed that the biochemical transformations in pea aphids were the same as those established in *Nepeta*. Then we compiled a list of candidates from the 96 genes specifically expressed in sexual female hind legs that also encoded enzymes that could, in principle, carry out these predicted reactions. One putative *IDS* gene, two putative phosphatase genes, one putative cytochrome P450 gene, one putative P450 reductase gene, and six putative oxidase/reductase genes were selected for further characterization (Fig. 1B).

**ApIDS Catalyzes the Metal Ion Cofactor–Dependent Formation of GPP.** The formation of GPP in the horseradish leaf beetle *Phaedon cockerelliae* is catalyzed by a bifunctional IDS (PcIDS) that shows a metal ion cofactor–dependent product specificity, producing primarily GPP with cobalt or manganese, or farnesyl diphasphate (FPP) with magnesium (12). A homolog of PcIDS (ApIDS, gene ID, 100144905) was specifically expressed in sexual female hind legs. Phylogenetic analysis revealed that PcIDS and ApIDS clustered together in a clade of beetle and aphid GPP/ FPP synthases (Fig. 2A). In vitro characterization of recombinant ApIDS (*SI Appendix*, Tables S3 and S4) showed IDS activity when incubated with IPP and DAMPP. Using magnesium as a cofactor, ApIDS produced similar amounts of GPP, FPP, and geranylgeranyl diphasphate (GGPP), while cobalt and, to a lesser extent, manganese shifted the product specificity to GPP as the major product (Fig. 2B and *SI Appendix*, Fig. S2). This suggests that ApIDS, like PcIDS, uses cobalt or manganese as a metal ion cofactor to produce GPP for iridoid formation in vivo.

**The Phosphatase ApGES Hydrolyzes GPP to Geraniol.** In plants, the formation of the mono- and sesquiterpene alcohols geraniol and farnesol from GPP and FPP is catalyzed by terpene synthases (TPSs) (23). However, in insects, farnesol, which is an intermediate in juvenile hormone biosynthesis, is produced from FPP by a phosphatase belonging to the haloalkanoic acid dehalogenase (HAD) super family (24, 25). Our candidate search in the pea aphid transcriptome did not reveal any TPS- or HAD-like proteins that were selectively expressed in sexual female hind legs. Instead, we identified two putative phosphatases annotated as dolichylidiphosphatase (gene ID, 100158803) and inositol polyphosphate-1-phosphatase (gene ID, 100162683) (Fig. 1B and *SI Appendix*, Tables S3 and S5). The putative inositol polyphosphate-1-phosphatase, expressed in *Escherichia coli* as a soluble protein, showed no GPP hydrolysis activity. In contrast, *Saccharomyces cerevisiae* microsomes harboring this recombinant membrane-bound putative dolichylidiphosphatase protein hydrolyzed GPP to geraniol (Fig. 2C and *SI Appendix*, Fig. S3). Thus, we named this dolichylidiphosphatase homolog ApGES.

The P450 ApG8H and the P450 Reductase ApRed Act Together to Produce 8-Hydroxygeraniol. Both plants and the leaf beetle *P. cockerelliae* utilize P450 enzymes to catalyze the hydroxylation of geraniol to 8-hydroxygeraniol (2, 9, 15). Only a single P450 in the pea aphid transcriptome displayed selective expression in sexual female hind legs, and this enzyme grouped together with PcG8H from *P. cockerelliae* in a phylogenetic analysis (Fig. 1B and *SI Appendix*, Fig. S4), though these two proteins share only 35% amino acid sequence identity. The aphid gene was named ApG8H and the complete open reading frame (ORF) was expressed in *S. cerevisiae* either alone or together with a putative P450 reductase gene (ApRed, gene ID, 100162683), which had a similar expression pattern to ApG8H. In the presence of the cosubstrate NADPH and ApRed, ApG8H converted geraniol to 8-hydroxygeraniol (Fig. 2D). A heterologously expressed P450 from maize (BX2) (26) that was used as a negative control, and ApG8H expressed without ApRed, showed no activity. Nota-bly, ApG8H exhibited a relatively broad substrate specificity, hydroxylating citronellol, nerol, linalool, and neral, though not the monoterpene hydrocarbons limonene and myrcene (*SI Appendix*, Fig. S5), suggesting that an oxygen atom at C1 is critical for binding or catalysis.

**The Short-Chain Reductase ApHGO Catalyzes the NADP+–Dependent Oxidation of 8-Hydroxygeraniol to the Iridoid Precursor 8-Oxogeranial.** While the oxidation of 8-hydroxygeraniol to 8-oxogeranial in plants is catalyzed by alcohol dehydrogenases (5, 9), *P. cockerelliae* beetles use a flavin-dependent glucose-methanol-cholin (GMC) oxidase to catalyze this reaction (13). Thus, we tested two aphid short-chain alcohol dehydrogenase (SDR) candidates, one annotated as farnesol dehydrogenase (gene ID, 100301633) and the other as retinol dehydrogenase (gene ID, 100162094), as well as two GMC candidates (gene IDs, 100169582 and 100164798) that were selectively expressed in sexual female hind legs (Fig. 1B and *SI Appendix*, Fig. S6). The complete ORFs were expressed in *E. coli* and purified proteins were assayed with 8-hydroxygeraniol in the presence of NADP+. Enzyme activity could only be observed for the putative farnesol dehydrogenase (named ApHGO), which catalyzed this two-step oxidation (Fig. 2E and *SI Appendix*, Fig. S7). Further characterization showed that ApHGO preferred NADP+ over NAD+ as cosubstrate and exhibited a broader substrate specifcity, also oxidizing geraniol and nerol, but not β-citronellol (*SI Appendix*, Fig. S8). The putative retinol dehydrogenase 100162094 and the GMC oxidase 100169582, although not able to accept 8-hydroxygeraniol as substrate, converted geraniol to geranial (*SI Appendix*, Fig. S7).

**The Iridoid Synthase ApISY Is a Membrane Protein Catalyzing the Reduction of 8-Oxogeranial.** The crucial formation of the cyclopentanoid-pyran scaffold occurs with the reductive cyclization
of 8-oxogeranial. In plants, this step is initiated by iridoid synthase, an SDR that belongs to the progesterone Sβ-reductase superfamily (3). The enzyme that insects use to catalyze this reduction is unknown. In vitro assays of a putative retinol dehydrogenase (gene ID, 100162094) and two GMC oxidases (gene IDs, 100169582 and 100164798) from our list of candidate genes showed no activity with the substrate 8-oxogeranial and the cosubstrate NADPH (SI Appendix, Table S5). However, a putative oxidoreductase annotated as membrane-bound polyphenol reductase (gene ID, 103310029) (SI Appendix, Table S5) was also selectively expressed in sexual female hind legs. Yeast microsomes containing this recombinant protein converted the substrate 8-oxogeranial to cis-trans-iridodial and five other unidentified compounds (Fig. 3A and SI Appendix, Figs. S9 and S10). Thus, the tested protein was named ApISY. As with plant iridoid synthases, without NADPH, ApISY showed no activity (Fig. 3A).

In Nepeta, and likely also in other plants, iridoid synthase works in concert with cyclases that mediate the stereoselective cyclization of the initial reduction product of ISY, 8-oxocitronellol enol/eolate, into different nepetalactol stereoisomers (8). When any plant ISY is assayed in vitro without a cyclase, the product profile is strongly dependent on the assay conditions. In high buffer concentrations or at low pH values, spontaneous tautomeration of the 8-oxocitronellol enol/eolate intermediate to 8-oxocitronellol is favored, while low buffer conditions or higher pH values lead to the spontaneous cyclization to cis-trans-nepetalactol as the predominant product. Moderate buffer concentrations or pH values lead to a mixture of monocyclic dialdehydes (8). Using ISY from the plant C. roseus (CrISY) as a point of comparison (Fig. 4 and SI Appendix, Fig. S11), we were surprised to observe that assays with yeast microsomes containing ApISY showed specificity for formation of cis-trans-nepetalactol over a broad buffer concentration range. Only at a buffer concentration of 0.5 M, yield of cis-trans-nepetalactol was affected in the ApISY reaction (Fig. 4).

Since the plant ISY is a soluble protein, while ApISY is an integral membrane protein with seven predicted transmembrane domains (SI Appendix, Table S5), we assayed CrISY in the presence of yeast
ApISY Likely Evolved from a Polyrenol Reductase Ancestor.

Polyrenol reductases, ubiquitous in eukaryotes, catalyze reduction of the α-isoprene unit of polyrenols to form dolichols, the precursors for dolichol-linked monosaccharides that are required for protein N-glycosylation (27–29). Together with steroid 5α-reductases and very-long-chain enoyl-CoA reductases, polyrenol reductases belong to the steroid 5α-reductase (SRD5A) family (Pfam, PF02544). A BLAST analysis with ApISY as query revealed two polyrenol reductase-like genes in most of the available aphid genomes. In a phylogenetic analysis, these genes formed two distinct and aphid-specific clades among the polyrenol reductases of eukaryotes (Fig. 3 B and SI Appendix, Fig. S12).

The Flavin-Dependent GMC Oxidase ApNEPO Converts cis-Trans-Nepetalactol into cis-Trans-Nepetalactone. The SDR NEPS1 from Nepeta catalyzes the oxidation of cis-trans-nepetalactol to cis-trans-nepetalactone (8). To identify the enzyme that catalyzes this reaction in pea aphids, the putative retinol dehydrogenase (gene ID, 100162094) and the two FAD-dependent GMC oxidases (gene IDs, 100169582 and 100164798) from our candidate gene list (Fig. 1 B) were assayed with the cis-trans-nepetalactol substrate and NADP+. While the SDR 100162094 and the GMC oxidase 100164798 were not active, GMC oxidase 100169582 converted cis-trans-nepetalactol into cis-trans-nepetalactone (Fig. 5 and SI Appendix, Fig. S13). A phylogenetic analysis revealed that this enzyme, designated A. pisum nepetalactol oxidase (ApNEPO), belonged to the ε-clade of GMC oxidoreductases (SI Appendix, Fig. S14) and was not related to the leaf beetle enzyme PcHGO, which clustered into a beetle-specific GMC clade (SI Appendix, Fig. S14). Sequence prediction suggested that ApNEPO contains a signal peptide targeting the protein into the lumen of the endoplasmic reticulum (ER) (SI Appendix, Table S5).

Discussion

Here, we elucidated the entire iridoid pathway in the pea aphid A. pisum. Previously reported feeding studies in pea aphids (21), as well as the spatial localization of nepetalactone biosynthesis (20), guided the identification of the six biosynthetic genes (Fig. 1 A). The characterization of these enzymes indicates that the plant and aphid nepetalactone biosynthetic pathways are composed of the same chemical transformations. However, each of the respective enzymes clearly evolved independently in plants and aphids. Our data show that in some cases plants and aphids recruited enzymes from different protein families to catalyze the

![Fig. 3](https://www.pnas.org/content/119/42/e2211254119.full/fig3)

**Fig. 3.** ApISY is a membrane-bound reductase that likely evolved from a polyrenol reductase ancestor. (A) Yeast (S. cerevisiae) microsomes made from a strain harboring the EV or expressing ApISY were assayed with 8-oxogeranial either in the presence or absence of NAD(P)H. Reaction products were extracted with ethyl acetate and analyzed using GC-MS. (B) Rooted tree of polyrenol reductases and iridoid synthases from Hemiptera. The tree was inferred by using the maximum likelihood method based on the Le_Gascuel_2008 model. Bootstrap values (n = 1,000) are shown next to the branches. A discrete gamma-distribution was used to model evolutionary rate differences among sites. The tree is drawn to scale, with branch lengths measured in the number of amino acid substitutions per site. All positions with less than 95% site coverage were eliminated. A putative polyrenol reductase from the Western flower thrips, Frankliniella occidentalis, was used to root the tree. NCBI accession numbers for all sequences are given in the tree.
some reactions (SI Appendix, Fig. S15). The formation of geraniol in Catharanthus and Nepeta, for example, is mediated by terpene synthases (5, 9), while the pea aphid uses a phosphatase to produce geraniol by direct hydrolysis of the phosphodiester bond of GPP (Fig. 2C). Moreover, the reductive cyclization of 8-oxogeranial to cis-trans-nepetalactol and the subsequent oxidation of this alcohol to cis-trans-nepetalactone in aphids involves the action of a polyprenil reductase-like protein and a flavin-dependent oxidase from the GMC family, respectively, while plants recruited members of the SDR family to catalyze both reactions (SI Appendix, Fig. S15).

Only one other natural product pathway, the three-step biosynthesis of the cyanogenic glycoside linamarin, has been fully elucidated in both plants and insects. Recent work has demonstrated that linamarin biosynthesis consists of two cytochrome P450s and one glucosyl transferase in both plants and insects, and that these pathways arose independently (30, 31). Additionally, terpene synthases, the key enzymes in terpene formation, have been identified in both plants and insects, and these enzymes are also the result of independent evolution in the different kingdoms (32–34).

Iridoids and iridoid-related compounds are widespread among insects and have been observed in different insect orders, including Coleoptera, Hymenoptera, and Hemiptera (1). The discovery of the aphid nepetalactone pathway provides an opportunity to determine whether iridoids evolved convergently in two divergent species of insects. The biosynthesis of the iridoid-related dialdehyde chrysomelidial has been partially elucidated in the leaf beetle P. cockleariae (11). Although chrysomelidial lacks the cyclopentanoid-pyran scaffold that defines the iridoids, its formation shares many of the same reactions as the iridoid core pathway in aphids (SI Appendix, Fig. S15). GPP is produced in the leaf beetle by PcIDS, which is obviously phylogenetically related to ApIDS (12) (Fig. 2A). Geraniol, which is produced in beetles by an as yet undiscovered enzyme, acts as substrate for PcG8H, a P450 that hydroxylates this alcohol to 8-hydroxygeraniol (15). Although PcG8H and ApG8H both belong to clan 3 of insect P450s (SI Appendix, Fig. S4), their low sequence identity suggests independent origins. Independent evolution of enzyme activities is even more obvious for the last two steps of the chrysomelidial pathway. The oxidation

**Fig. 4.** Buffer concentration and the lipid environment influence iridoid synthase activity. (A) Yeast (S. cerevisiae) microsomes containing ApISY were assayed with 8-oxogeranial in the presence of NADPH under different buffer concentrations. (B) CrISY was expressed in E. coli, purified, and assayed with 8-oxogeranial and NADPH under different buffer concentrations in the background of yeast microsomes (YM). (C) Purified CrISY was incubated with 8-oxogeranial and NADPH under different buffer concentrations. Reaction products were extracted with ethyl acetate and analyzed using gas chromatography–mass spectrometry. 1, cis-trans-iridodial; 2, unidentified; 3, unidentified; 4, cis-trans-nepetalactol; 5, unidentified; 6, unidentified; 7, unidentified; 8, trans-trans-iridodial; 9, cis-trans-iridodial; 10, tetrahydro-8-oxogeranial; 11, unidentified; and 12, unidentified.

**Fig. 5.** Biochemical characterization of ApNEPO. ApNEPO and another putative GMC-oxidase (ApGMC-Ox) highly expressed in hind legs of A. pism sexual females were expressed as N-terminal His-tag fusion proteins in E. coli, purified, and incubated with 7S-cis-trans-nepetalactone in the presence of NADP. Enzyme products were extracted with ethyl acetate and analyzed using gas chromatography–mass spectrometry.
of 8-hydroxygeraniol in *P. cochiniae* is catalyzed by a GMC oxidase (13), in contrast to aphids, which recruited an SDR for the same reaction, and the final nonreductive cyclization of the formed 8-oxogeranial to chrysomelidial is presumably mediated by a Pcto-like juvenile hormone-binding protein, although conclusive evidence of this enzyme activity is still lacking (35). Overall, the elucidation of the iridoid pathway in aphids presented here shows that although the reaction sequence is conserved, iridoid formation evolved independently not only in different kingdoms, but also in different insect orders through convergent evolution (*SI Appendix*, Fig. S15).

From a chemical perspective, the first committed step in the iridoid pathway, the cyclization of 8-oxogeranial by ISY, is of mechanistic interest. All known plant iridoid synthases belong to the SDR protein family and have been described to catalyze the reduction of 8-oxogeranial to a highly reactive 8-oxocitronellyl enol/enolate intermediate, which is then cyclized by NEPS or MLPL proteins to different stereoisomers of nepetalactol (8, 9). In the absence of a cyclase, 8-oxocitronellyl enol/enolate can react spontaneously to form various compounds depending on the assay conditions. Higher pH values or low buffer concentrations lead to cyclization to cis-trans-nepetalactol, while acidic conditions or high buffer concentrations favor spontaneous formation of 8-oxocitronellal or other dialdehydes (8). In contrast to plant ISY, a soluble protein likely located in the cytosol, the pea aphid ISY is an integral membrane protein (8, 10). A protein with a signal peptide in *A. pisum* genome with a sequence similarity to ISY was described in *A. pisum* (10). We cannot rule out that ApISY itself or other, as yet unidentified, aphid proteins fulfill this function in vivo. Most aphid species produce cis-trans-nepetalactol (18, 36). The damson-hop aphid *Phorodon humili*, however, produces the cis-cis isomer (37, 38). This species must have either an iridoid synthase with different stereoselectivity, catalyzing both the reduction and cyclization of 8-oxogeranial to the final nepetalactol stereoisomer, or a partner cyclase that cyclizes the potential 8-oxocitronellyl enol/enolate to cis-trans-nepetalactol, presumably by preventing contact of the reactive intermediate with proton donors or other acidic compounds. Since the lipid composition of membranes can vary significantly depending on cell type and developmental stage, a detailed lipid analysis may help to better understand this process. Although iridoid formation in pea aphids does not appear to require the action of a cyclase, we cannot rule out that ApISY itself or other, as yet unidentified, aphid proteins fulfill this function in vivo. Most aphid species produce cis-trans-nepetalactol (18, 36). The damson-hop aphid *Phorodon humili*, however, produces the cis-cis isomer (37, 38). This species must have either an iridoid synthase with different stereoselectivity, catalyzing both the reduction and cyclization of 8-oxogeranial to the final nepetalactol stereoisomer, or a partner cyclase that cyclizes the potential 8-oxocitronellyl enol/enolate intermediate to cis-cis-nepetalactol. Elucidating the biosynthetic pathway for cis-cis-nepetalactol in *P. humili* would provide additional insight into the complex chemistry underlying the formation of the iridoid backbone in animals.

Sequence comparisons revealed that ApISY is related to polypre-nol reductases, a class of integral membrane proteins involved in *N*-glycosylation of secreted and membrane-bound proteins (27–29) (SI Appendix, Fig. S12). Interestingly, most of the aphid species sequenced to date possess two putative polypre-nol reductase copies that form two distinct and aphid-specific clades in a phylogenetic tree of the polypre-nol reductases of Hemiptera (Fig. 3B and SI Appendix, Fig. S12). Thus, it is likely that the ApISY-containing clade represents aphid iridoid synthases, while the other clade contains true polypre-nol reductases, which is a metabolic enzyme that is essential for survival. The close relationship of these two clades suggests that iridoid synthase activity arose by gene duplication and subsequent neofunctionalization of a polypre-nol reductase gene early in aphid evolution or in an ancestor of the aphids. Furthermore, the striking sequence similarities among proteins within the two clades (Fig. 3B) indicates a high degree of purifying selection to preserve their respective enzymatic functions. We cannot predict the evolutionary origin of other iridoid pathway genes because the function of their closest homologs in aphids is still unknown. For example, although ApGES was annotated as a dolichylphosphatephosphatase, an enzyme acting together with polypre-nol reductase in the process of *N*-glycosylation of secreted and membrane-bound proteins (39, 40), there is no experimental evidence for dolichylphosphatephosphatase activity of ApGES, and BLAST analysis did not reveal any other *ApGES*-related gene that could be dedicated to dolichylphosphatephosphatase function in the *A. pisum* genome.

A defining feature of the iridoid biosynthetic pathway in aphids is that many of the enzymes appear to be membrane anchored. In addition to the iridoid synthase ApISY, the phosphatase ApGES, the cytochrome P450 ApG8H, and its reductase ApRed were predicted to be integral membrane proteins (SI Appendix, Table S5). Moreover, the prediction of an ER signal peptide of ApNEPO suggests that this protein is localized in the lumen of the ER (SI Appendix, Table S5). This leads us to speculate that these enzymes may form a metabolon, most likely, based on signal sequence prediction, on the membrane of the ER (Fig. 6). Given the predicted mitochondrial signal peptide of ApIDS, the pathway likely starts in the mitochondria with the formation of GPP, which is then transported to the ER. Notably, the list of candidate iridoid genes exclusively expressed in hind legs of sexual female aphids contained seven genes annotated as transporters for the inner and outer membrane of mitochondria that could be involved in GPP transport (SI Appendix, Table S1). The hydrolysis of GPP and the formation of the iridoid backbone could then be catalyzed by the putative metabolon in the ER membrane, which may be a metabolon that could be dedicated to dolichylphosphatephosphatase function in the *A. pisum* genome.

![Fig. 6. The iridoid pathway in aphids might be organized in a membrane-associated metabolon. While ApHGO is predicted to be localized as soluble protein in the cytosol and ApNEPO is likely localized in the lumen of the ER, ApGES and ApISY are transmembrane proteins.](https://doi.org/10.1073/pnas.2211254119)
provide efficient substrate channeling, preventing the release of highly reactive pathway intermediates such as 8-oxoguanine. The final oxidation of the comparatively stable nepeta lactol product to nepetalactone could then occur in the lumen of the ER, and formation of nepetalactone-containing vesicles and their transport to the cell membrane might represent a possible mechanism for the active release of these volatile iridosides.

Overall, chemical logic, along with the discreet spatial localization of the site of biosynthesis, facilitated the discovery of the six-step pathway for nepetalactone biosynthesis in animals. This provides a foundation for understanding how complex natural products have evolved in two kingdoms of life. The insect pathway also provides insights into the relatively understudied field of insect natural product biosynthesis.

### Materials and Methods

**Cultivation of A. pismum and Generation of Sexual Female Aphids.** Asexual females of pea aphid (*A. pismum* [Harri]) clone JML06 were reared on 4-wk-old broad bean (*V. faba*) cv. “The Sutton” plants under long-day conditions (16 h light/dark, 22 °C, 60% humidity). To avoid escape of aphids, plants were covered with air-permeable cellophane bags (18 × 38.5 cm, Griesinger Verpackungsmittel). To generate sexual female and male aphids, asexual L3 aphid larvae were transferred to short-day conditions mimicking fall season (12 h light/dark, 22 °C, 45% humidity). Two generations later, sexual females and males were produced, and adult aphids (6 to 10 d old) were used for experiments. The emission of iridosides by sexual female aphids was tested by placing a solid phase microextraction (SPME) fiber for 3 h into the headspace of *V. faba* plants with the aphids. The SPME fiber was then loaded into the injector of a gas chromatograph coupled with a mass spectrometer (GC-MS) as described below.

**Transcriptome Sequencing and Gene Identification.** For RNA extraction, 20 hind legs of sexual females, 20 front legs of sexual females, 20 hind legs of male aphids, and 20 hind legs of asexual aphids were collected and directly placed in 450 μL lysis buffer containing guanidinium thiocyanate (imuPreP RNA Mini Kit, IST Innsscreen). Material was shredded by shaking with metal beads using a Tissue Lyzer II (Qiagen) for 2 × 4 min (frequency 50%). Total RNA was extracted with the imuPreP RNA Mini Kit according to the manufacturer’s instructions, eluted in 30 μL RNase-free water, and sent to Novogene for RNA-seq library construction (polyA enrichment) and sequencing (NovaSeq PE150, paired reads, 6 gigabytes of raw data per sample). Trimming of the obtained sequencing reads and mapping to the pea aphid genome (version 3) were performed with the program CLC Genomics Workbench (Qiagen Bioinformatics) (mapping parameter: length fraction, 0.8; similarity fraction, 0.9; maximum number of hits, 25). In order to identify pea aphid genes involved in iridoid formation, we performed Pearson correlation based on the hypothesis that iridoid for-

**Prediction of Signal Peptides and Transmembrane Domains.** Prediction of signal peptides and transmembrane domains (SI Appendix, Table S1) was performed using TargetP-2.0 (https://services.healthtech.dtu.dk/service.php?TargetP-2.0) and DeepTMHMM (https://dtu.biolib.com/DeepTMHMM/), respectively.

**Gene Synthesis and Cloning.** The complete ORFs of *ApHGO*, *ApNEPO*, *ApIDS*, and *ApISY* were expressed using *E. coli* cells, as listed in SI Appendix, Table S6. *ApG8H* and *ApRed* were cloned as sticky-end fragments into the same pESC-Leu-2d vector using the two different cloning sites (41). *ApSY* and *ApGES* were separately cloned as sticky-end fragments into pESC-Leu-2d. CDNA was synthesized from total RNA (1 μg) treated with DNasel (Thermo Fisher Scientific) using SuperScript III reverse transcriptase and oligo (dT)20 primers (Invitrogen) according to the manufacturer’s instructions. All synthesized or amplified sequences are given in SI Appendix, Table S4.

**Heterologous Expression of Candidate Genes in *E. coli.*** Expression constructs were transformed into *E. coli* strain BL21 (DE3) (Invitrogen). Liquid cultures were grown in lysogeny broth at 37 °C and 220 rpm until an OD_{600} of 0.7, induced with a final concentration of 0.5 mM isopropyl-beta-D-thiogalactopyranoside, and subsequently incubated at 18 °C and 220 rpm for 16 h. The cells were harvested by centrifugation at 3,200 × g for 10 min, resuspended in refrigerated extraction buffer (50 mM Tris-HCl pH 8.5, 500 mM NaCl, 20 mM imidazole, 5% [vol/vol] glycerol, 50 mM glycine, ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (1 tablet/50 mL buffer, freshly added), and lysosome (10 mg/50 mL buffer, freshly added) and disrupted by sonication for 2 min (2 s on, 3 s off) on ice (Bandelin UW 2070). Cell debris were removed by centrifugation (35,000 × g at 4 °C for 20 min) and the N-terminal His-tagged proteins were purified from the supernatant using NiNTA agarose (Qiagen) according to the manufacturer’s instructions. The buffer of the eluted protein samples was exchanged for assay buffer (for details see paragraph enzyme assays) 100 mM 3-(N-morpholino)propanesulfonic acid (MOPS) pH 7.5, 10% (vol/vol) glycerol by using Amicon 10K Concentrator columns (Merck Millipore). Sodium dodecyl sulfate-polyacrylamid gel electrophoresis and spectrophotometric analysis was used to check purity and approximate quantity of proteins.

**Heterologous Expression of ApGES, ApG8H, and ApISY in *S. cerevisiae.*** For heterologous expression in yeast, constructs were transformed into the *S. cer-

**Enzyme Assays.** ID5 assays were carried out using 3 μg of purified ApIDS protein desalted in ID5 assay buffer (25 mM MOPS, pH 7.2, 10% [vol/vol] glycerol, 1 mM metal ion cofactor (MgCl₂, MnCl₂, or CoCl₂), and the substrates IPP and DMAPP (each 50 μM) in 100 μL assay buffer at 30 °C for 1 h. Product formation was monitored by liquid chromatography-tandem mass spectrometry (LC/MS/MS) as described below.
GES activity was determined in assays (total volume, 100 μL) containing 20 μL yeast microsomes harboring ApGES, 25 mM Tris-HCl (pH 7.5) and 50 μg/μL GPP. The assays were overlaid with 100 μL hexane and incubated for 20 min at 22 °C. Enzyme products were extracted by vortexing for 1 min, and 1 μL of the hexane phase was injected into the GC-MS (see below). For measuring G8H activity, 10 μL microsomes harboring either ApG8H alone or in combination with the P450 reductase ApRed were incubated in 25 mM sodium phosphate buffer (pH 7.0) with 25 mM substrate (geraniol, geraniol, citronellol, or citronellol, respectively) and 1 mM NADPH in a total volume of 100 μL for 2 h at 30 °C. Assays were then overlaid with 100 μL ethyl acetate and vortexed for 1 min. G8H products were analyzed by injecting 1 μL of the ethyl acetate phase into the GC-MS. The indole hydroxylase BX2 that is involved in benzoazinoid formation in maize (26) was used as a negative control. Screening of ApG8H activity with other substrates including cital A+8, nerol, linalool, limonenone, and myrcene was performed by adding 10 μL of the substrate (0.5 mM dissolved in methanol) to 500 μL of living yeast cells induced with galactose-containing medium (see above). Cells were further incubated for 24 h at 28 °C and 200 rpm, and afterward extracted with 200 μL ethyl acetate. An aliquot (1 μL) of the organic phase was injected into GC-MS for enzyme product analysis.

HGO activity was analyzed using assays containing 40 μg/μL purified protein, 1 mM NAD+ or NADP+, respectively, and 0.5 mM substrate (β-hydroxygeraniol, geraniol, nerol, or β-citronellol) in a total volume of 50 μL MOPS buffer (0.1 M). Assays were overlaid with 200 μL ethyl acetate, incubated for 2 h at 30 °C, and enzyme products were extracted by vortexing the assay for 1 min. One microliter of the organic phase was injected into GC-MS for enzyme product analysis.

ISY activity was measured in assays (total volume, 50 μL) containing 20 μL microsomes, 50 mM MOPS pH 7.5, 1 mM NADPH, and 0.5 mM 8-oxogeraniol. Assays were incubated for 2 h at 30 °C, overlaid with 100 μL ethyl acetate, and products were extracted by vortexing the assays for 1 min. One microliter of recombinant and purified CrISY from C. roseus (8) was tested under the same conditions as described above either in the presence or absence of 20 μL yeast microsomes as negative control.

NEPO activity was determined as described above for HGO with 3 μg of purified protein and 0.5 mM 7cis-trans-nepetalactone as substrate.

Gas Chromatography–Mass Spectrometry Analysis. Qualitative analysis of volatile sex pheromones released from sexual female pea aphids was conducted using an Agilent 6890 Series gas chromatograph coupled to an Agilent 5973 mass spectrometer with a carrier gas (He) at 25 m/s, overlayed with 100 μL yeast microsomes, 50 mM MOPS pH 7.5, 1 mM NADPH, and 0.5 mM 8-oxogeraniol. Assays were incubated for 2 h at 30 °C, overlaid with 100 μL ethyl acetate, and products were extracted by vortexing the assays for 1 min. One microliter of recombinant and purified CrISY from C. roseus (8) was tested under the same conditions as described above either in the presence or absence of 20 μL yeast microsomes as negative control.

Data, Materials, and Software Availability. Raw reads from the transcriptome sequencing were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the BioProject accession PRJNA6664370 (44). Amplified gene sequences were deposited in NCBI GenBank with the accessions ON862918 (ApG8H), ON862919 (ApRed), and ON862920 (ApISY) (47). All other study data are included in the article and/or SI Appendix.

ACKNOWLEDGMENTS. We thank Michael Reichelt and Maritta Kunert for help with LC-MS/MS and GC-MS analysis, respectively; Nestor Jose Hernandez-Lozada for providing purified recombinant CrISY protein; Kimberly Falk for drawing the aphid cartoons shown in Fig. 1; Julie Jaquery for helpful advice regarding the generation of sexual female aphids; and Elke Goschala, Claudia Pakebusch, and Andreas Webers from the greenhouse team of the Max Planck Institute for Chemical Ecology for rearing V. faba plants. We gratefully acknowledge Eden Fei Li for preliminary experiments. This work was supported by the Max Planck Society. Additional support was provided by the European Research Council (788301).