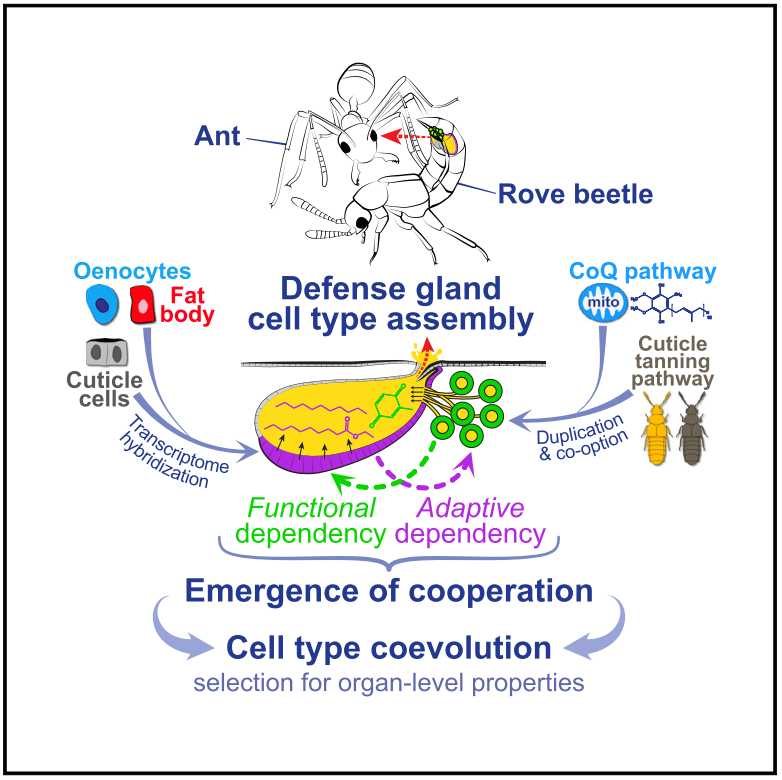
Cell

Evolutionary assembly of cooperating cell types in an animal chemical defense system

Graphical abstract



Highlights

- A rove beetle defense gland reveals how evolution builds new organs
- Two secretory cell types work together, one making a solid toxin, the second a solvent
- New cellular functions evolved by combining ancient transcriptome modules and pathways
- Evolution of each cell type was shaped by adaptive coevolution between the cell types

Authors

Adrian Brückner, Jean M. Badroos, Robert W. Learsch, Mina Yousefelahiyeh, Sheila A. Kitchen, Joseph Parker

Correspondence

joep@caltech.edu

In brief

The rove beetle defense gland is the result of coevolution between two secretory cell types that work together to produce a chemical cocktail, illuminating how cell types evolve cooperative interactions generating organ-level behaviors.

Brückner et al., 2021, Cell *184*, 6138–6156
December 9, 2021
https://doi.org/10.1016/j.cell.2021.11.014

CellPress

CellPress

Cell

Article

Evolutionary assembly of cooperating cell types in an animal chemical defense system

Adrian Brückner,[1] Jean M. Badroos,[2] Robert W. Learsch,[3] Mina Yousefelahiyeh,[1] Sheila A. Kitchen,[1] and Joseph Parker[1,4,*]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[2]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125, USA
[3]Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA
[4]Lead contact
*Correspondence: joep@caltech.edu
https://doi.org/10.1016/j.cell.2021.11.014

SUMMARY

How the functions of multicellular organs emerge from the underlying evolution of cell types is poorly understood. We deconstructed evolution of an organ novelty: a rove beetle gland that secretes a defensive cocktail. We show how gland function arose via assembly of two cell types that manufacture distinct compounds. One cell type, comprising a chemical reservoir within the abdomen, produces alkane and ester compounds. We demonstrate that this cell type is a hybrid of cuticle cells and ancient pheromone and adipocyte-like cells, executing its function via a mosaic of enzymes from each parental cell type. The second cell type synthesizes benzoquinones using a chimera of conserved cellular energy and cuticle formation pathways. We show that evolution of each cell type was shaped by coevolution between the two cell types, yielding a potent secretion that confers adaptive value. Our findings illustrate how cooperation between cell types arises, generating new, organ-level behaviors.

INTRODUCTION

Animal diversification has been shaped by the evolution of new organs with novel functional properties (Raff, 1996). Structures such as eyes, skeletons and brains were multicellular innovations that opened new ecological frontiers and catalyzed major cladogenesis (Erwin, 2021). How the functions of complex organs originate has challenged biologists since Darwin (Darwin, 1859; Dawkins, 1986; Lynch, 2007; Oakley and Speiser, 2012; Shubin et al., 2009; West-Eberhard, 2003). Organs are composed of specialized cell types that perform dedicated roles within these structures (Schmidt-Rhaesa, 2007); yet, knowledge of how cell types are constructed at the molecular level during evolution, and how they cooperate to produce emergent, organ-level behaviors, remains fragmentary. At the gene regulatory level, new cell types are thought to arise from the evolution of unique, combinatorial transcription factor activities that specify novel cell identities (Arendt et al., 2016). However, a major problem is explaining how the downstream molecular components, which carry out the function of the differentiated cell, themselves originate. For a cell to execute its role, multiple gene products must be configured to work in concert. In a multicellular organ, this process must be coordinated across multiple cell types, fostering a division of labor that generates organ behavior (Kishi and Parker, 2021).

The molecular changes leading to new cell type- and organ-level functions are obscured by the deep antiquity of many animal organs (Shubin et al., 2009). However, insight may be gained from "evolutionary novelties"—lineage-specific structures that are absent in outgroups or serially homologous body parts (Erwin, 2015; Wagner and Lynch, 2010). Vertebrate eyes (Lamb et al., 2007), insect wings (Linz and Tomoyasu, 2018), beetle horns (Hu et al., 2019) and the mammalian placenta (Griffith and Wagner, 2017) are paradigms that connect genomic and developmental changes to the establishment of new phenotypes. Nevertheless, knowledge of how the functions of novel organs arise at the cellular level has been limited. One hurdle has been the challenge of measuring the transcriptomic composition of cell types—critical for tracing how different subfunctions within a cell evolved. The advent of single-cell methods has now brought many questions regarding cell-type evolution within reach (Marioni and Arendt, 2017). New tools permit decomposition of the transcriptome, enabling discovery of gene expression programs underlying cell identity and quantitative assessment of relationships between cell types (Way and Greene, 2019). Here, we leverage these methods to trace the evolution of cooperating cell types in an animal organ.

Exocrine glands are archetypal organ novelties: unique, clade-specific glands have evolved convergently thousands of times in metazoans, transforming how animals interact via pheromones and chemical defenses, enabling specialized modes of prey capture and digestion, and aiding in substrate adhesion, desiccation avoidance, and antimicrobial protection (Brückner and Parker, 2020). Each origin of a new gland involves the evolution of new secretory cell types that synthesize natural products (Torres

and Schmidt, 2019). Venom glands of wasps, scorpions, and cone snails; the explosive weaponry of bombardier beetles; and diverse mammalian scent glands are multicellular novelties composed of taxon-restricted cell types found nowhere else. For many such organs, including vertebrate salivary glands (Roussa, 2011), moth silk glands (Suzuki et al., 1990), cnidarian digestive glands (Babonis et al., 2019), and venom glands of many species (Surm and Moran, 2021), the final secretion is a cocktail produced by distinct secretory cell types. Biosynthetic synergism between cell types is pronounced in many small molecule-based chemical defense systems. Here, an inactive toxin and its activating enzyme may be produced by separate glands; alternatively, a soluble toxin and its solvent are secreted into a common reservoir, yielding a bioactive secretion (Kishi and Parker, 2021). Such functional interdependence encapsulates how emergent, organ-level behaviors can arise from cooperation between cell types.

One animal clade notable for pervasive glandular innovation is the rove beetles (Staphylinidae)—a radiation of 64,000 predominantly soil-dwelling predators that comprise Metazoa's largest family (Betz et al., 2018; Parker, 2017). Rove beetles exhibit widespread evolution of abdominal glands that produce diverse compounds including quinones, terpenes, alkaloids, iridoids, hydrocarbons and esters (Francke and Dettner, 2005). A unique, flexible body enables rove beetles to target secretions at other organisms, fostering interspecies relationships from chemical defense to symbiosis (Parker, 2016). The glands are often composite organs that produce multi-compound secretions via distinct cell types (Dettner, 1993a). Although their chemistry and anatomy have been studied in many species (Francke and Dettner, 2005), how these organs function at the molecular level, and how they evolved in specific lineages, is unknown. Here, we examine the molecular architecture of one such gland, enabling us to infer how a novel organ and its constituent cell types evolved. We exploited the model species *Dalotia coriaria* of the megadiverse Aleocharinae, a subfamily that possess a defensive "tergal gland" on the dorsal abdomen (Steidle and Dettner, 1993). Using single-cell transcriptomics, we uncover pathways for defensive compounds in tergal gland cell types. By inferring transcriptomic and pathway relationships with more ancient cell types, we retrace how the tergal gland was functionally assembled during evolution. We present evidence that evolution followed an accessible path toward cell-type cooperativity, building an organ capable of producing a multi-compound secretion that confers adaptive value.

RESULTS

The tergal gland is an organ novelty

When provoked, *Dalotia* flexes its abdomen to release the contents of a large chemical defense gland—the tergal gland—situated between abdominal segments A6 and A7 (Figures 1A and 1B) (Steidle and Dettner, 1993). The secretion is smeared directly onto the perceived threat and contains three benzoquinones (BQs) (Figure 1D)—aromatics that are common arthropod defensive compounds (Blum, 1981). BQs are topical irritants that bind TRPA1 channels (Ibarra and Blair, 2013), activating nociceptive pathways (Kang et al., 2010). The adaptive value of *Dalotia*'s secretion is demonstrated by experimentally depleting the contents of the gland, disarming the beetle, and compromising its defense against predatory ants (Figures 1C and S1A). The BQs may also have antimicrobial properties (Carcamo-Noriega et al., 2019).

Although BQs are highly noxious, each compound in isolation is a solid. Consequently, the secretion also contains four less toxic compounds: large amounts of a C11 alkane, *n*-undecane, low amounts of two C10 esters, ethyl decanoate and isopropyl decanoate, and trace levels of a C12 ester, ethyl dodecanoate (Figure 1D). These compounds may function as solvents, combining with BQs to create a potent secretion, as well as promoting spreading and penetration into cuticles (Blum, 1981; Dettner, 1984). The secretion is thus a cocktail, the composition of which may be explained by possible synergism between its components. To produce this cocktail, the tergal gland consists of two anatomically distinct cell types: (1) a secretory epithelium, composed of columnar cells that are continuous with the epidermis and form a large reservoir inside the body into which they secrete directly (Figures 1E and 1F); and (2) large, bulbous secretory cells, 8–14 per animal, located posterior to the reservoir into which they feed via tubular ducts (Figures 1E and 1F). The tergal gland, and these two cell types, are specific to aleocharines (Steidle and Dettner, 1993); they are absent in outgroup rove beetles (subfamily Tachyporinae) that cannot produce defensive compounds (Figures 1G, S1B, and S1C). We sought to determine how these two cell types cooperate to synthesize *Dalotia*'s defensive secretion.

The solvent cell type

Dalotia's ability to produce defensive compounds is refractory to antibiotics (Figures S1D–S1H). We therefore reasoned that biosynthesis relies not on symbiotic bacteria, as in some other chemically defended rove beetles (Piel, 2002), but on enzymes encoded in the beetle's genome and expressed within the gland. We dissected small numbers of each gland cell type and performed SMART-Seq RNA sequencing, creating cell type transcriptomes (Figures S2A–S2C). This approach led to recovery of candidate enzymes in each cell type. We focused initially on synthesis of the undecane and esters, hypothesizing that their hydrocarbon chains derive from fatty acids (Stanley-Samuelson et al., 1988). Fatty acids are built from acetyl-CoA and malonyl-CoA units produced by glycolysis (Wakil et al., 1983). Rearing *Dalotia* on D-glucose-$^{13}C_6$ food, we observed strong ^{13}C incorporation into the undecane and esters (Figure S3A), confirming they originate via the glycolytic pathway and fatty acid synthesis. In contrast, adding D_{23}-dodecanoic acid to food led to negligible D_{23} incorporation (Figure S3B). The compounds thus do not derive from consumed fatty acids.

In animals, the multistep reaction to produce fatty acids is catalyzed by a multidomain fatty acid synthase (FASN) (Smith, 1994; Smith et al., 2003). FASNs generally synthesize the commonest fatty acids, C16 and C18 in length (Smith et al., 2003). However, *Dalotia*'s gland compounds are shorter (C10–C12). Notably, a predicted *FASN* was upregulated in the epithelial secretory cells comprising the reservoir (Figure S2B) but not in the bulbous secretory cells (Figure S2C). Strong expression in these cells was confirmed via *in situ* hybridization chain reaction

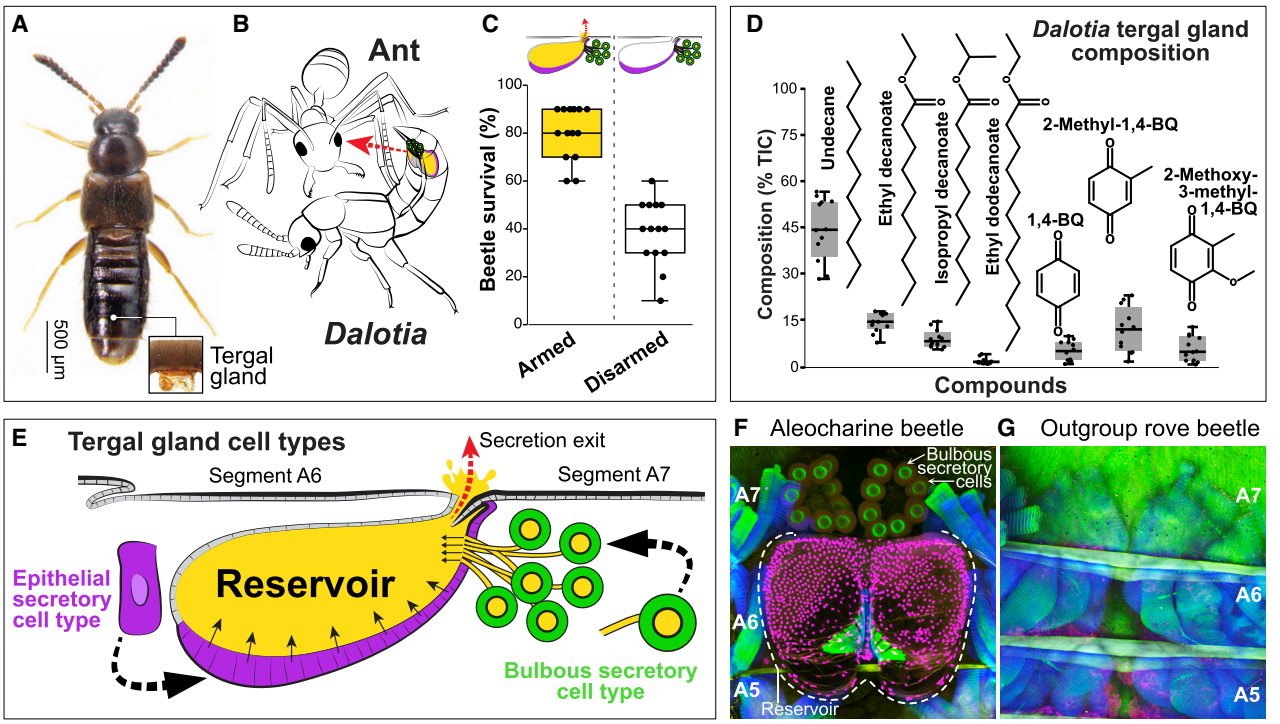


Figure 1. The tergal gland

(A) *Dalotia coriaria* with position of tergal gland.

(B) Cartoon of gland deployment.

(C) *Dalotia* survival with predatory *Liometopum* ants after 48 h (armed versus disarmed glands; Mann-Whitney U test: $p < 0.001$). The middle bar of each boxplot represents the median, the box denotes the interquartile range, while the whiskers show the minimum and maximum.

(D) Composition of gland secretion (n = 10 beetles) measured by GC-MS. The middle bar of each boxplot represents the median, the box denotes the interquartile range, while the whiskers show the minimum and maximum.

(E) Cartoon cross section of abdomen showing gland cell types: secretory epithelium (magenta) invaginated between segments A6–A7 forms a chemical reservoir; bulbous secretory cells (green) connect to reservoir via ducts. The secretion exudes via the opening between segments.

(F) Confocal image of internal segments A5–A7 of aleocharine abdomen (magenta, α-Engrailed antibody labels reservoir cell nuclei; green, bulbous secretory cells autofluorescence; blue, phalloidin-stained muscle).

(G) Confocal image of internal segments A5–A7 of glandless outgroup rove beetle (Tachyporinae: *Coproporus*).

See also Figures S1 and S2 and Table S1.

(HCR) (Choi et al., 2018) (Figure 2F). To deduce this *FASN*'s function, we temporally silenced it with systemic RNA interference (RNAi) (Parker et al., 2018; Tomoyasu and Denell, 2004) (Figure S2D). Silencing caused near-total loss of undecane and all three esters without impacting the BQs (Figures 2A, 2B, and S3C). We name this enzyme "Master Fatty Acid Synthase" (MFASN), due to its upstream role in synthesizing fatty acid-derived compounds in the tergal gland. We infer that the epithelial secretory cells (magenta cell type in Figure 2J) are the source of these compounds. This cell type is hence named the "solvent cell type."

Further analysis of the solvent cell transcriptome led to discovery of a complete fatty acid pathway downstream of MFASN. The maximally expressed transcript is a predicted cytochrome P450 (CYP) (Figure S2B), cell-type-specific expression of which was confirmed by HCR (Figure 2G). Phylogenetic analysis revealed that this CYP belongs to an insect-specific "CYP4G" class (Figure S3G), known to decarbonylate aldehydes to create alkanes (Feyereisen, 2020; Qiu et al., 2012). Consistent with this function, silencing the tergal gland *CYP4G* (herein "*TG-CYP4G*") caused selective loss of the undecane and the appearance of its C12 aldehyde precursor, dodecanal (Figures 2C and S3D). We infer that TG-CYP4G functions downstream of MFASN as a terminal enzyme that decarbonylates dodecanal to make undecane. To produce dodecanal, the activated C12 fatty acid (dodecanoyl-CoA) produced by MFASN must first be reduced to an aldehyde. Conspicuously, a predicted fatty acyl-CoA reductase (herein "*TG-FAR*") was strongly expressed in solvent cells (Figures 2H and S2B). As with *TG-CYP4G*, silencing *TG-FAR* caused specific loss of undecane, this time without the appearance of dodecanal (Figures 2D and S3E). We infer that solvent cells make undecane via a pathway in which MFASN produces dodecanoic acid that is reduced by TG-FAR and decarbonylated by TG-CYP4G (Figure 2J).

In addition to this alkane pathway, a second branch of the solvent pathway yields the esters. Silencing *MFASN* caused loss of the three esters (Figure 2B). Because two of these are C10 compounds (isopropyl and ethyl decanoate), MFASN must produce

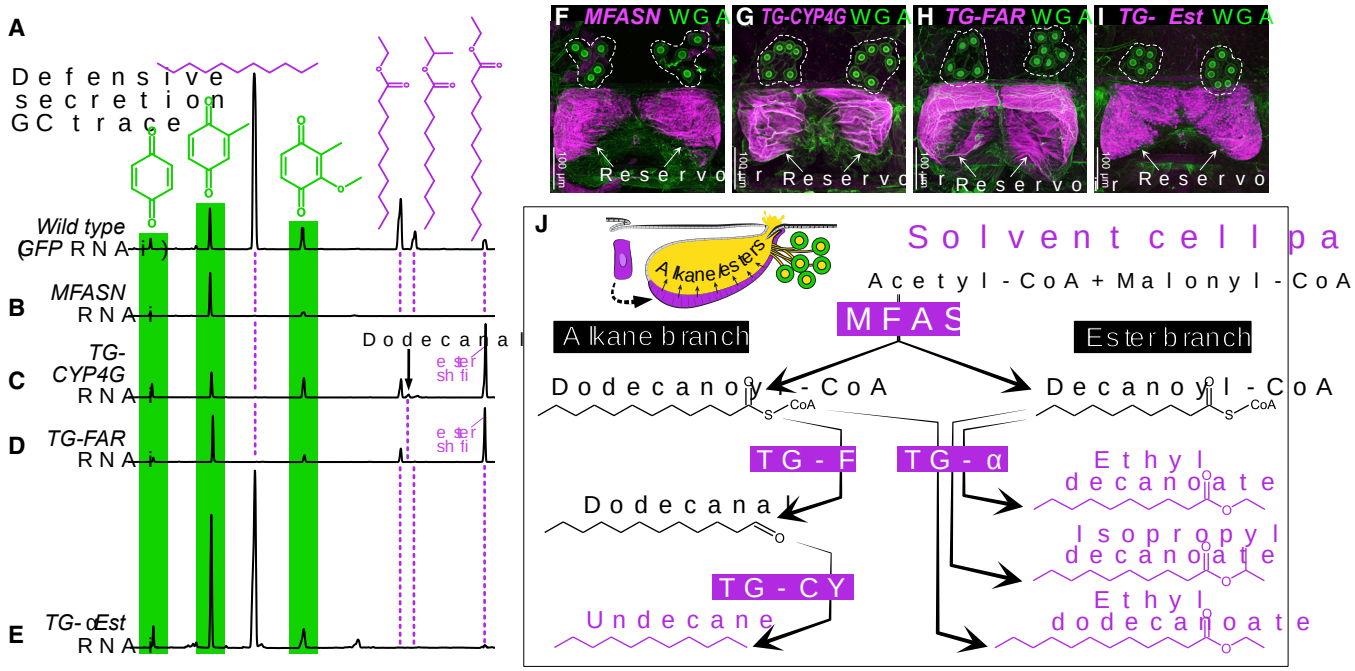


Figure 2. The solvent cell type

(A–E) Tergal gland secretion GC traces.

(A) Wild-type with compound peaks labeled.

(B) *MFASN* RNAi causes loss of undecane and esters but not BQs.

(C) *TG-CYP4G* RNAi removes undecane; dodecanal appears and C12 ester is strongly increased.

(D) *TG-FAR* RNAi removes undecane; dodecanal is absent and the C12 ester increases.

(E) *TG-αEst* RNAi removes esters without affecting undecane or BQs. For statistics related to (B)–(E), see Table S1.

(F–I) HCR (magenta) of solvent pathway enzymes: *MFASN* (F), *TG-CYP4G* (G), *TG-FAR* (H), and *TG-αEst* (I). Green, WGA.

(J) The solvent pathway.

See also Figures S1 and S3.

their C10 fatty acid precursor, decanoic acid, in addition to the dodecanoic acid precursor of undecane (Figure 2J). Notably, we observed strong upregulation of a single α-esterase enzyme in solvent cells (Figures 2I and S2B). Silencing this enzyme ("*TG-αEst*") removed all three esters without affecting undecane (Figures 2E and S3F). TG-αEst is thus the key enzyme in the ester branch of the solvent pathway. TG-αEst esterifies both C10-CoA and C12-CoA to yield the three esters (Figure 2J).

A cell atlas of *Dalotia* abdominal segments

Solvent cells create a volatile solution that we hypothesize dissolves the BQs. To investigate how this function evolved, we reconstructed the evolutionary steps by which the solvent cells were pieced together. We applied single-cell RNA sequencing (scRNA-seq) using droplet microfluidics (10x Chromium v.3) to entire adult abdominal segments, enabling us to determine the transcriptomic relationships of solvent cells to other cell types in the *Dalotia* abdomen. We created two segmental atlases: "Segment 7" comprising segment A7 plus the solvent cell reservoir (Figure 3A); and "Segment 6" encompassing segment A6 minus the reservoir. Segment 6 is serially homologous but lacks the gland, approximating ancestral segment morphology (Figures 3A and 3B). We integrated five 10x runs (Figures S4A–S4C) and selected 3,000 transcripts that capture transcriptional heterogeneity across cells for subsequent analyses. Using Seurat 3.1 (Butler et al., 2018), we determined 26 cell clusters, assigning these to putative cell types using marker genes from *Drosophila* (see STAR Methods). The atlas is shown in Figure 3C as a UMAP.

Using each cell type's mean expression of the 3,000 transcript set, we inferred a tree of transcriptomic similarity, and used the topology to test whether the 26 cell types could be grouped into higher-order classes via a random forest classifier (Figures S4D and S4E). This classifier produced strong support for 14 higher groupings, herein called "cell classes" (Figures 3C, S4E, and S4F). Overall, 85% of all cells were correctly assigned to one cell class (Figure S4E). The 26 cell types and 14 cell classes thus represent a robust classification of cell types in abdominal segments 6 and 7 of *Dalotia*.

Ester branch evolution via co-option of fat body enzymes

Within the abdominal cell atlas, a single cluster co-expressed the four solvent pathway enzymes (Figures 3D–3G and S5A–S5F). This cell type—cluster number 6—is thus the solvent cell type, a determination further supported it being almost exclusively restricted to segment 7 (Figure 3B). To investigate solvent pathway evolution, we searched for features shared with other

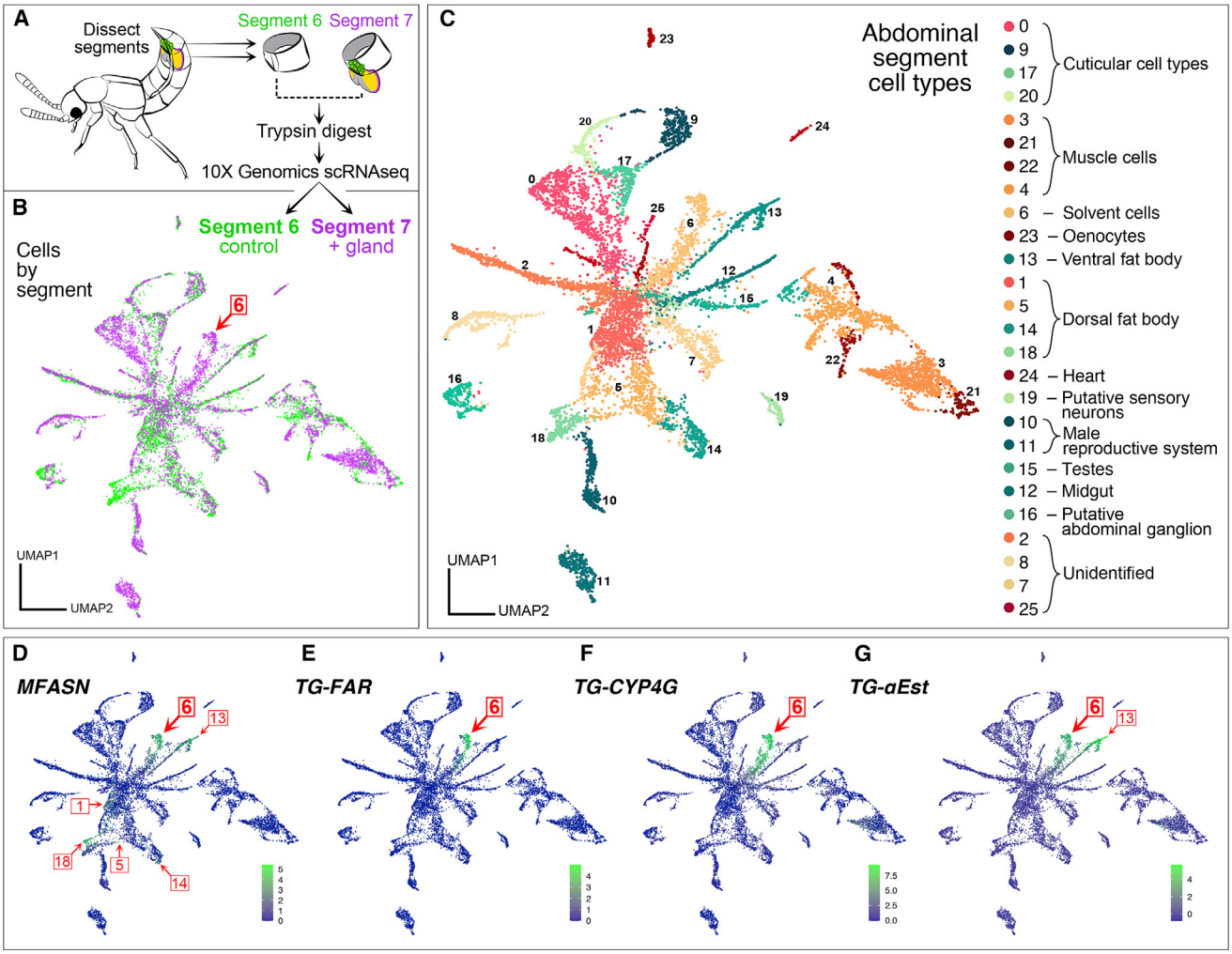


Figure 3. Single cell atlas of *Dalotia* abdominal segments

(A) Scheme for 10x scRNA-seq.

(B) UMAP of cells separated by segment (green, Segment 6; magenta, Segment 7). Cell type 6 is composed primarily of Segment 7 cells.

(C) UMAP of cell types and their grouping into cell classes.

(D–G) Expression of solvent pathway enzymes: *MFASN* (D), *TG-FAR* (E), *TG-CYP4G* (F), and *TG-αEst* (G). Numbers correspond to cell types in (C).

See also Figure S4.

cell types. We observed that several cell types—1, 5, 13, 14, and 18—expressed *MFASN* to high levels, indicating fatty acid synthesis (Figures 3D and S5A). We deduced that these cell types comprise *Dalotia*'s fat body—an ancient insect organ sharing many functions with vertebrate liver and adipose tissue (Arrese and Soulages, 2010). Fat body cells are sites of fat and glucose storage, as well as lipid synthesis, and are identifiable as lipid droplet-containing cells forming loose tissue throughout the abdomen. HCR revealed strong *MFASN* expression in *Dalotia*'s fat body (Figure S5G). Phylogenetic analysis shows that *MFASN* is a single copy gene present in all beetles, as well as in *Drosophila* (Figure 4L). *MFASN*'s deep conservation and broad expression across *Dalotia*'s fat body cell types imply an ancestral function in the insect fat body. We HCR-labeled *MFASN* in a distantly related outgroup beetle, *Tribolium* (Tenebrionidae), revealing strong, fat body-specific expression (Figure S5H). *MFASN*'s ancestral function was indeed in the fat body; production of C12 and C10 precursors evolved by co-opting *MFASN* into solvent cells.

Fat body cells can be highly heterogenous in insects (Haunerland and Shirk, 1995), and this is so in *Dalotia*. First, cell types 1, 5, 14 and 18—but not 13—cluster-based on mean transcript expression (Figure S4D), implying two different fat body classes. The four former cell types also express *Drosophila* fat body markers including *pumpless* and *apolipophorin*, whereas cell type 13 does not (Figures 5I–5I″). Most strikingly, the two fat body classes differ in expression of α-esterases (Data S1): all α-esterase copies except one are expressed in cell types 1, 5, 14, and 18 (and occasional other cell types) (Data S1), whereas cell type 13 expresses only *TG-αEst*, shared exclusively with

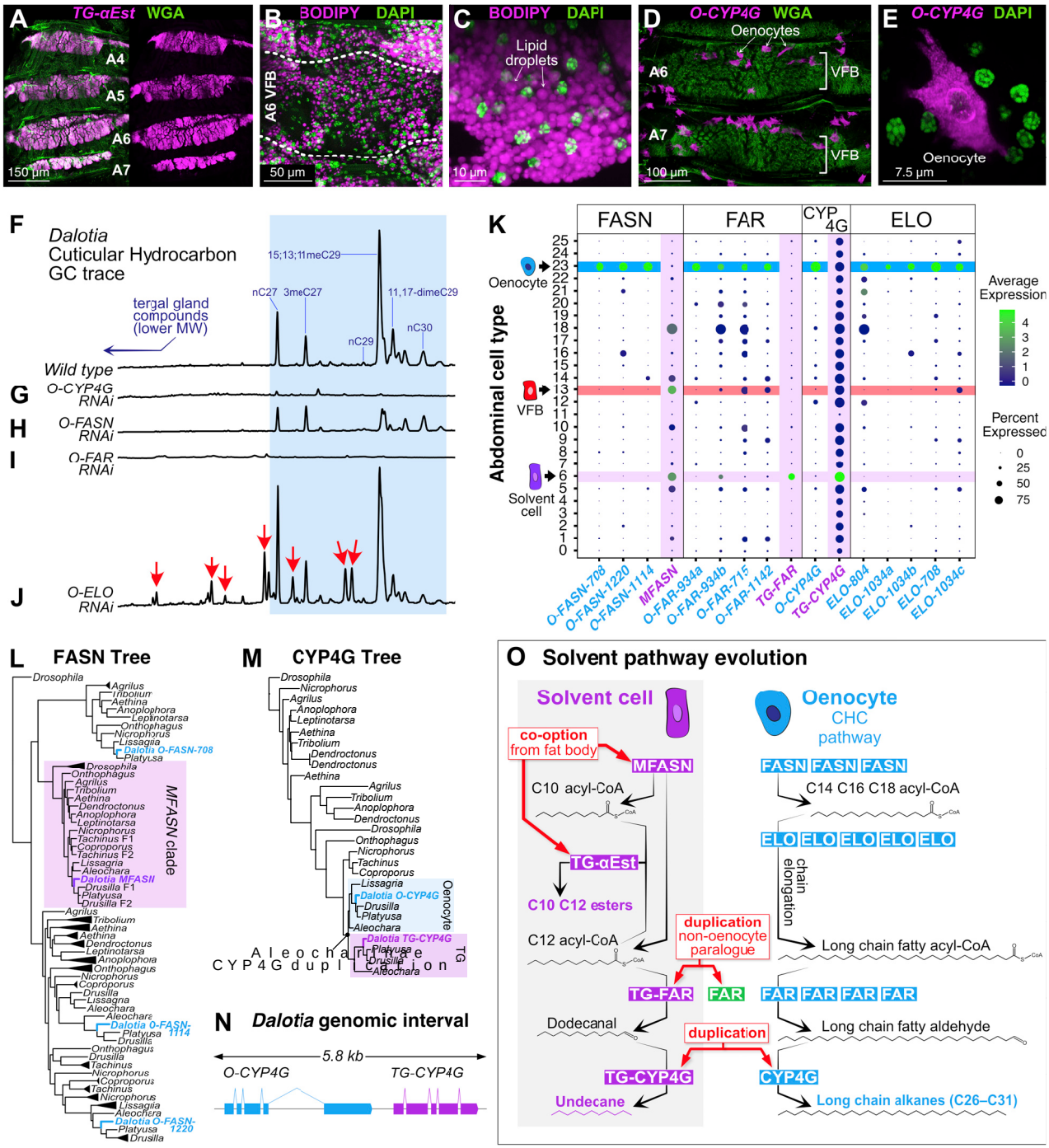


Figure 4. Solvent pathway evolution

(A) *TG-αEst* HCR (magenta) labels segmental bands of ventral fat body (VFB; green, WGA-stained membranes).

(B and C) BODIPY (magenta) labels lipid droplets in VFB (green, DAPI).

(D and E) *O-CYP4G* HCR (magenta) stains oenocytes in ventral abdomen. (D) Intermingling of oenocytes and VFB (green, WGA). (E) Single oenocyte with *O-CYP4G* expression (green, DAPI).

(F–J) GC traces of *Dalotia* CHCs. (F) Wild-type with major peaks identified. (G) *O-CYP4G* RNAi removes CHCs. (H) *O-FASN* RNAi removes most CHCs. (I) *O-FAR* RNAi removes CHCs. (J) *O-ELO* RNAi shifts CHCs to shorter chain lengths.

(K) Expression dot plot of *FASN*, *FAR*, *CYP4G* and *ELO* enzyme copies across abdominal cell types. Dot size, percent of cells of each type expressing transcript; color, mean expression level.

(legend continued on next page)

solvent cells (Figure 3G; Data S1). Using HCR of *TG-αEst*, we identified cell type 13 as subepidermal bands of fat body containing large lipid droplets (Figures 4A–4C). We refer to cell type 13 as "ventral fat body" (VFB) and other fat body types distributed more dorsally as "dorsal fat body" (DFB; cell types 1, 5, 14+18) (Figures S4D and S5I). Given the general function of α-esterases in the fat body, that sister paralogs of *TG-αEst* are expressed primarily in fat body, and that *TG-αEst* itself functions in the VFB, we infer that the solvent pathway's ester branch evolved by recruiting a fat body α-esterase to function downstream of MFASN (Figure 4O). Importantly, *TG-αEst* is an aleocharine-specific gene that is absent in other beetles, including outgroup Tachyporinae (Data S1). Furthermore, the duplicate occurs in genomes of species known to produce esters, including *Dalotia* (tribe Athetini), *Drusilla*, and *Platyusa* (both Lomechusini), but is absent from *Aleochara* (Aleocharini) that does not (Data S1).

Alkane branch evolution via duplication of oenocyte enzymes

Many aleocharines synthesize esters as the putative BQ solvent but lack alkanes, which may represent a more recent evolutionarily addition (Steidle and Dettner, 1993). Recruitment of MFASN and TG-αEst may thus have been an early step in solvent pathway evolution, with the alkane branch added subsequently. We noted that one cell type—23—expressed a paralog of *TG-CYP4G* (Figure 4K; Data S2A). CYP4Gs have been shown to function canonically in oenocytes—ancient pheromone-producing cells found in all insects (Makki et al., 2014). Oenocytes synthesize very long-chain (C25–C40) alkanes and alkenes known as cuticular hydrocarbons (CHCs) that are secreted onto the body and encode a chemical signature of species identity (Billeter et al., 2009). CHCs also form a waxy barrier against desiccation (Koto et al., 2019; Qiu et al., 2012). In oenocytes, CYP4G performs terminal decarbonylation to yield the secreted CHCs (Qiu et al., 2012). We HCR-labeled this *CYP4G*, revealing clusters of enlarged cells subepidermally in the ventral abdomen, often intermingled with VFB (Figures 4D and 4E). Silencing this locus led to near-total loss of *Dalotia*'s CHCs (Figures 4F and 4G; Data S2E), confirming that these cells are the oenocytes. We name this enzyme *O-CYP4G* (*Oenocyte-CYP4G*).

Recent studies have delineated a conserved oenocyte CHC pathway (Blomquist and Ginzel, 2021). This pathway bears striking similarity to the alkane branch of the solvent pathway: oenocytes use FASNs to produce fatty acids that are reduced to aldehydes by FARs before decarbonylation by CYP4G (Blomquist and Ginzel, 2021; Holze et al., 2021)—modifications that mirror exactly the steps for undecane synthesis (Figure 2J). Consistent with the deep conservation of the insect CHC pathway, we find that, in addition to *O-CYP4G*, cell type 23 expresses three *FASN* paralogs and four *FAR*s (Figure 4K; Data S2B and S2C). As with *O-CYP4G*, silencing selected copies of these enzymes strongly diminished CHC production (Figures 4H and 4I; Data S2F and S2G). Hence, *Dalotia* expresses parallel alkane pathways in ancient oenocytes and novel solvent cells, their close similarity implying common ancestry. Indeed, we find that *TG-CYP4G* and *O-CYP4G* are sister duplicates found only in Aleocharinae (Figure 4M); in *Dalotia*, the two copies sit tandemly in the genome (Figure 4N). The ancestral CYP4G was likely an oenocyte enzyme that synthesized CHCs; duplication gave rise to oenocyte and solvent cell copies (Figure 4O). Indeed, *TG-CYP4G* is conserved in species that produce alkanes (*Aleochara*, *Drusilla*, *Platyusa*, and *Dalotia*) but is absent from *Lissagria* that does not (Figure 4M). Whether duplication of an *O-FAR* led to *TG-FAR* is less clear; *FAR*s undergo extensive gene birth-and-death, so their history is challenging to infer (Finet et al., 2019). *TG-FAR* is not a sister paralog of an *O-FAR* (Data S2J), so its role in alkane synthesis may have arisen convergently. Regardless, the CHC pathway represented a pre-existing template: by recruiting a FAR and CYP4G downstream of MFASN, alkanes were added to the defensive secretion (Figure 4O).

Despite similar enzyme logic, the products of the two pathways are markedly different: the CHC pathway produces very-long-chain waxy hydrocarbons, whereas the solvent pathway makes a medium-chain volatile liquid. Oenocytes make longer compounds via elongases (ELOs) (Blomquist and Ginzel, 2021; Holze et al., 2021) (Figure 4O). We find *Dalotia* oenocytes express five ELOs (Figure 4K; Data S2D), knockdown of which caused an altered profile with many shorter chain compounds (Figure 4J; Data S2I) without reducing total CHC levels (Data S2H). In contrast, solvent cells express no ELOs (Figure 4K; Data S2D). Hence, the solvent pathway evolved via selective FAR and CYP4G recruitment without an ELO, enabling medium-chain biosynthesis (Figure 4O).

Solvent cell evolution via transcriptomic hybridization of ancient cell types

Solvent cells comprise part of the beetle's cuticle, forming part of the intersegmental membrane joining segments A6 and A7 (Figure 1E). Developmentally, the cells derive from the epidermal posterior (P) compartment of A6 and express Engrailed (Figure 1F; Data S2K and S2L), the P compartment selector transcription factor (Morata and Lawrence, 1975). Like the surrounding epidermis, solvent cells produce chitin, which forms an internal lining to the reservoir and is continuous with the rest of the exoskeleton (Steidle and Dettner, 1993). Despite their epidermal identity, solvent cells are enlarged, columnar secretory cells that produce undecane and esters. To understand how solvent cells acquired this property, we studied their transcriptomic relationship to other cell types within the abdomen. We employed consensus non-negative matrix factorization (cNMF) (Kotliar et al., 2019), performing an

(L) ML tree of *FASN* genes from *Dalotia*, other beetles and *Drosophila* (LG+F+R6 model). Collapsed clades signify multiple paralogs. *Drusilla* and *Coproporus MFASN F1* and *F2* are fragments, likely of the same gene.
(M) ML tree of *CYP4G* genes (LG+I+G4 model). Duplication in Aleocharinae yields putative oenocyte and tergal gland (TG) clades.
(N) *Dalotia* genomic interval showing *TG-CYP4G* and *O-CYP4G* tandem copies.
(O) Model of solvent pathway assembly via co-option or duplication of VFB and oenocyte enzymes.
See also Figure S5, Table S1, and Data S1 and S2.

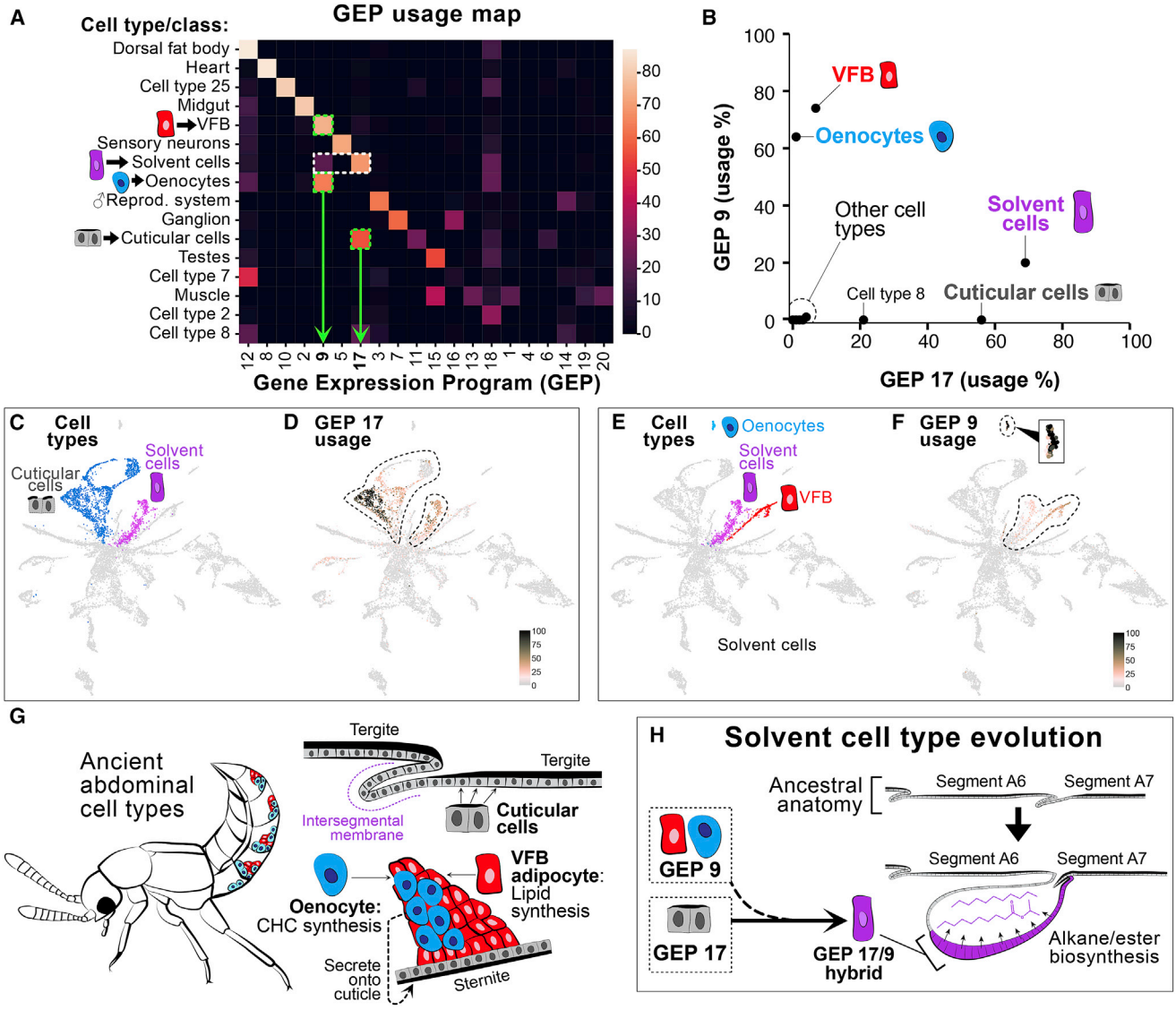


Figure 5. Solvent cell evolution

(A) GEP usage by cell type. Some cell types were grouped into classes for clarity and to facilitate computation.

(B) GEP9 and 17 usage by each cell type reveals solvent cells are a GEP9/17 hybrid.

(C) UMAP of cell types (from Figure 3C) with cuticle cell class and solvent cells highlighted.

(D) GEP17 usage across cell types with cuticle cells and solvent cells indicated.

(E) UMAP with oenocytes, VFB, and solvent cells highlighted.

(F) GEP9 across cell types with oenocytes, VFB and solvent cells indicated (inset is magnified oenocyte cluster).

(G) Cartoon showing segmentally repeated VFB and oenocytes in ventral abdomen.

(H) Model of solvent cell type evolution: cuticle cells comprising intersegmental membrane between A6 and A7 express GEP17 as principal GEP; acquisition of GEP9 from oenocytes and VFB transformed them into solvent cells.

See also Figure S5 and Data S3 and S4.

unsupervised search for constellations of significantly co-expressed transcripts across all cells within the 10x dataset, irrespective of their cell type. cNMF applies iterative NMF treatments on a transcripts-by-transcripts matrix to identify groups of significantly co-expressed genes ("gene expression programs" or "GEPs"). GEPs discretize the transcriptome into building blocks that may be surrogates of cellular properties. A GEP used in one or a few cell types may confer aspects of cell identity; conversely, a GEP used by many cell types likely underlies a routine activity such as mitosis (Kotliar et al., 2019). Employing cNMF, we determined that 20 GEPs accurately capture a decomposed representation of the total transcriptome of A6 and A7 (Data S3). We calculated the proportional contribution of each GEP to each individual cell type's

transcriptome, depicted as a usage map (Figure 5A). Visualizing transcriptome composition in this way, some GEPs appear cell-type-specific, such as GEP8 and GEP2—unique identity GEPs expressed in heart and midgut cells, respectively. Conversely, most cell types express GEPs 18 and 12, implying common cellular activities.

Of particular interest are cell types composed of combinations of GEPs (Figure 5A). These include the solvent cells, recovered as a composite of GEPs 17 and 9 (Figures 5A and 5B). Consistent with solvent cells being part of the cuticular epidermis, GEP17 is the principal GEP expressed by the cuticular cell class (Figures 5B–5D). This cell class is composed of cell types 0, 9, 17, and 20 (Figures 3C and S4D), which express multiple cuticle proteins and Laccase 2 that functions in cuticle tanning (Arakane et al., 2005) (Figure S6F). However, solvent cells are clearly divergent from other cuticular cells in their additional expression of GEP9, which accounts for ~20% of the transcriptome (Figure 5B). The composition of GEP9 reveals that it is constituted by 61 core transcripts (applying a stringent Z score filter of <0.002) (Data S4A) that show strong and significant enrichment for biological processes related to lipid metabolism and fatty acid biosynthesis (Data S4B–S4D). GEP9 is thus a transcriptional module that may endow solvent cells with their capacity for high level fatty acid production and modification.

Remarkably, GEP9 is the principal GEP of both oenocytes and VFB (Figures 5A, 5B, 5E, and 5F). This expression module therefore defines the two cell types from which solvent pathway enzymes were co-opted or duplicated. Unlike solvent cells, VFB and oenocytes show no pronounced GEP usage beyond GEP9 (aside from GEPs 18 and 12 that most cell types express) (Figure 5A). GEP9 thus likely contributes to the functional identity of these two cell types, which are both specialized for fatty acid biosynthesis. It follows that GEP9 probably imparts this same function in solvent cells. A close relationship between solvent cells, VFB, and oenocytes is further supported by the three cell types forming a clade based on mean transcript expression (Figure S4D), and random forest classifies them into the same cell class (Figures S4E and S4F). That solvent cells are a novelty within *Dalotia*'s cuticle, whereas oenocytes and fat body cells are ancient, non-cuticular cell types in all insects, implies a specific evolutionary scenario (Figures 5G and 5H). We suggest that solvent cells arose via transcriptomic hybridization: they are a cuticular cell type, ancestrally comprising intersegmental membrane, that gained an expression module that evolved in oenocytes and fat body, equipping them for high-level fatty acid synthesis (Figure 5H). As part of this process, the oenocytes and fat body also contributed distinct branches of the solvent pathway (Figure 4O).

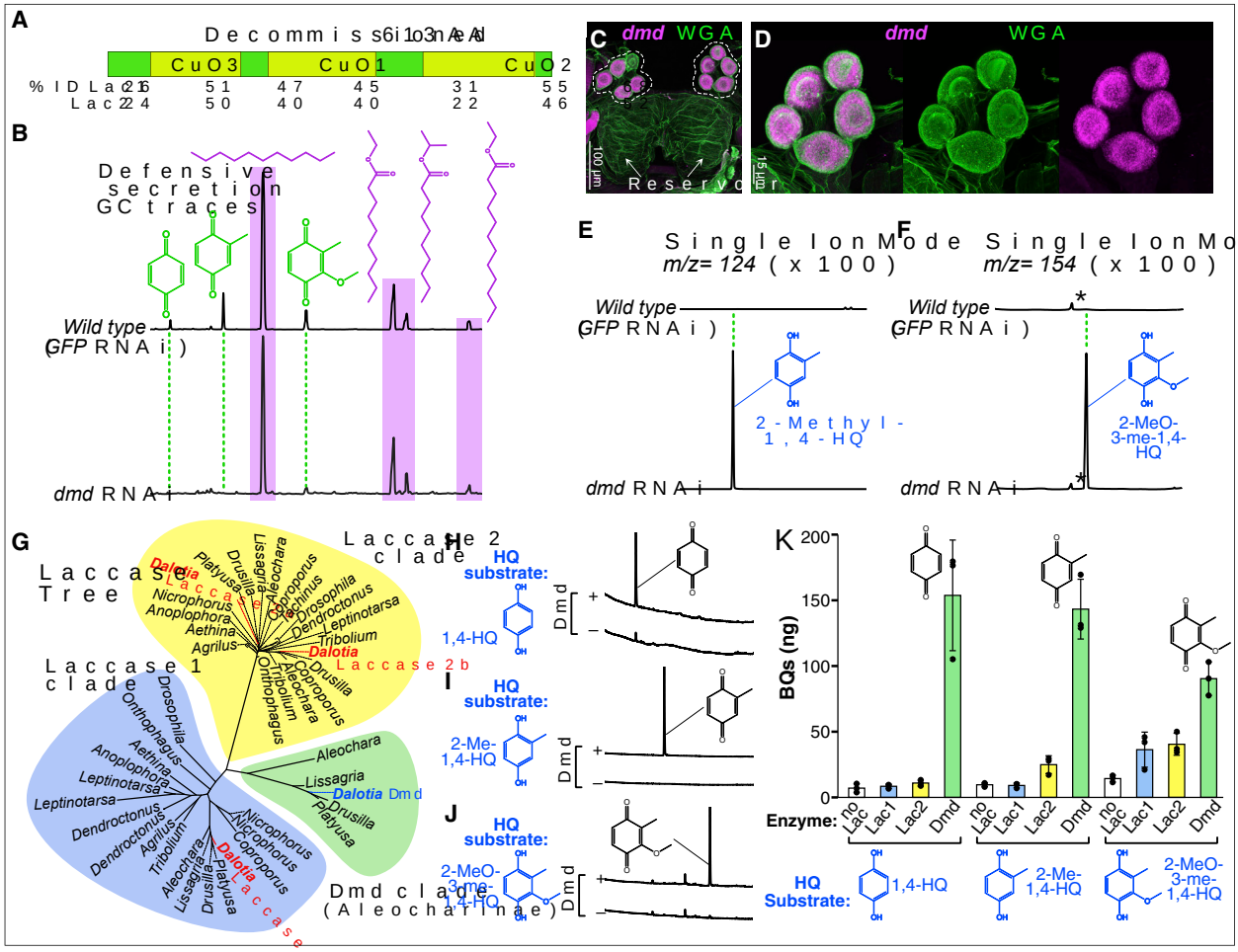
The BQ cell type

We next focused on BQ synthesis. Although defensive BQs have arisen in myriapods, arachnids, earwigs, termites, cockroaches, grasshoppers, capsid bugs, and at least seven times in beetles (Wagner et al., 2020), their mechanistic origins were hitherto unknown in any species. Aromatic compounds in animals are often acquired from dietary aromatic amino acids or symbiotic microbes, or sometimes synthesized *de novo* (Brückner et al., 2020; Torres et al., 2020). To infer the synthetic route, we fed *Dalotia* D-glucose-$^{13}C_6$ and observed negligible ^{13}C incorporation into the BQs (Figure S6A), arguing against complete *de novo* synthesis. Conversely, feeding Tyr-$^{13}C_6$ or Phe-$^{13}C_6$ led to strong ^{13}C incorporation, with molecular weights of all BQs increasing by exactly 6 (Figure S6A). The BQ's benzene rings thus derive from dietary aromatic amino acids.

Beyond their use in chemical defense, quinones play key roles in insect metabolism: ubiquinone (coenzyme Q_{10}) is a redox-active compound synthesized in mitochondria where it functions in electron transport (Stefely and Pagliarini, 2017); additionally, quinone intermediates arise during exoskeleton maturation (cuticle tanning) (Noh et al., 2016). In both contexts, aromatic amino acids are the precursors. We asked whether these ancient pathways could give clues to BQ synthesis in the tergal gland. In cuticle tanning, oxidation of Tyr-derived dopa and dopamine creates quinones that are pigment precursors and protein cross-linkers used for cuticle hardening (sclerotization) (Noh et al., 2016). Oxidation is mediated by Laccase 2 (Lac2), a secreted multicopper oxidase (MCO) (Asano et al., 2019). A predicted MCO was strongly upregulated in the second cell type within the gland—the bulbous secretory cells (Figure S2C). This transcript encodes a secreted protein with three cupredoxin domains in the same configuration as other laccases (Figure 6A) (Dwivedi et al., 2011). Silencing this laccase caused near-total loss of all three BQs, without affecting the solvents (Figures 6B and S6B). We name this laccase "Decommissioned" (Dmd) after its loss-of-function phenotype where the irritant is eliminated from the secretion.

We confirmed *dmd*'s expression in the bulbous secretory cells via HCR (Figures 6C and 6D), establishing the source of the BQs. We refer to these cells as the "BQ cell type." Because Dmd is secreted, we hypothesized that it may be a terminal pathway enzyme that oxidizes secreted BQ precursors. Laccases are well known to oxidize hydroxyl groups, and studies in other aleocharine rove beetles have shown that in addition to BQs, trace levels of corresponding hydroquinones (HQs) can occasionally be detected in the secretion, consistent with HQs being unoxidized precursors (Steidle and Dettner, 1993). In *Dalotia*, we sometimes recover trace levels of 2-methyl-1,4-HQ in the secretion, corresponding to the non-oxidized form of the highest abundance BQ. Strikingly, we find that although silencing *dmd* diminishes levels of 2-methyl-1,4-BQ, it leads to excess levels of 2-methyl-1,4-HQ (Figure 6E). Further, a HQ precursor of a different BQ species, 2-methoxy-3-methyl-1,4-HQ, additionally appears (Figure 6F). Accumulation of HQs following *dmd* silencing provides strong evidence that this laccase oxidizes secreted HQs, converting them into BQs. We relate this function to the BQ cell's anatomy, where the cell envelopes a lumen connected directly to the duct (Figure 6P). We posit that the lumen is an oxidation chamber into which Dmd and HQs are secreted and combine. The duct channels the resulting cytotoxic BQs into the solvent reservoir (Figure 6P).

We recovered *dmd* orthologs only in genomes of aleocharines that possess a tergal gland and synthesize BQs (Figure 6G). Phylogenetically, Dmd does not branch from inside either clade of conserved insect laccase, Lac1 and Lac2 (Figure 6G). As in other insects, *Dalotia lac2* is expressed in cuticular cells (Figure S6F) and silencing it abolishes tanning (Figures S6C and



(legend on next page)

S6D), whereas *dmd* is not expressed in the cuticle (Figure S6G) and is not involved in tanning (Figures S6C and S6E). Dmd thus defines an aleocharine-specific laccase clade that functions as a terminal oxidase in BQ synthesis. To corroborate this model, we synthesized Dmd protein and tested its ability to oxidize HQ precursors of all three BQs *in vitro*. Consistent with its inferred function, Dmd strongly catalyzed conversion of all three HQs to BQs (Figures 6H–6J and S6H). Further, we synthesized *Dalotia* Lac1 and Lac2 and found that although both enzymes exhibited some activity on at least one HQ substrate, neither were as efficient as Dmd (Figure 6K). We conclude that Dmd is a catalytically specialized laccase that performs HQ oxidation in BQ cells.

HQ synthesis via evolution of a ubiquinone-like pathway

What pathway produces the HQ precursors for Dmd? Our data indicate that BQs derive from Tyr (Figure S6A), but how the aromatic ring is hydroxylated to make HQs and decorated with methyl and methoxy groups is unknown. One eukaryotic pathway exists that integrates these steps: ubiquinone biosynthesis. Here, Tyr is converted to 4-hydroxybenzoic acid (4-HB), which is modified in the mitochondrion to yield an HQ, ubiquinol (Wang and Hekimi, 2019). *Dalotia*'s BQs resemble less modified versions of ubiquinol, at most incorporating methyl and methoxy groups (Figure 6L). To test whether HQs derive from a similar pathway, we fed *Dalotia* 4-HB-$^{13}C_6$ and measured $^{13}C_6$ incorporation into the BQs. Although the absolute level of incorporation was not high, we observed significant $^{13}C_6$ enrichment into the two most abundant BQs (Figure 6M). The magnitude of $^{13}C_6$ incorporation was lower than observed on feeding *Dalotia* Tyr-$^{13}C_6$ (Figure S6A), likely because 4-HB is a catabolite, not a dietary precursor, limiting its access from the gut to the correct cellular location. This result nevertheless identifies 4-HB as an intermediate in the conversion of Tyr to HQs.

Key steps in 4-HB's conversion to ubiquinol are carried out by sequentially acting CoQ enzymes (Wang and Hekimi, 2019) (Figure S7J). Two CoQ enzymes are of potential relevance: CoQ5, a methyltransferase that adds the methyl group, and CoQ3, an O-methyltransferase that creates the methoxy groups. We asked whether these enzymes add methyl or methoxy groups to *Dalotia*'s BQs. In most eukaryotes, including insects, CoQ enzymes are encoded by conserved single copy genes (Kawamukai, 2016). Due to their essential role in cellular respiration, studying them *in vivo* is challenging. For example, we observed ~3-fold higher transcription of *CoQ5* in BQ cells (Figure S7B), but knockdown with even low dsRNA levels led to complete lethality. Unusually, however, we found that *CoQ3* has duplicated in *Dalotia*, as well as in genomes of all other aleocharines surveyed (in some cases it has duplicated twice) (Figure 7F). Conspicuously, one of *Dalotia*'s *CoQ3*s is strongly upregulated in BQ cells (Figure S7D) as well as moderately so in putative sensory neurons (Figure S7H). Studies across eukaryotes have shown that CoQ3 adds two methoxy groups to ubiquinol using *S*-adenosylmethionine as the methyl donor (Poon et al., 1999). An analogous reaction might yield the methoxy group of 2-methoxy-3-methyl-1,4-BQ. We fed *Dalotia* CD_3-labeled methionine and observed direct incorporation of CD_3 into 2-methoxy-3-methyl-1,4-BQ, confirming that the same reaction takes place (Figure S7A). Remarkably, silencing the BQ cell-expressed *CoQ3* duplicate led to complete loss of 2-methoxy-3-methyl-1,4-BQ in the secretion (Figures 6O and S7E–S7G), without affecting *Dalotia*'s other BQs (Figures 6O and S7G). We name this *CoQ3* paralog "*methoxyless*" (*meos*) and deduce that it performs an identical modification to a defensive BQ as canonical CoQ3 does to ubiquinone. *meos* appears to be a single copy gene that is specific to aleocharines (Figure 6N).

Canonical CoQ3 functions on the mitochondrial inner membrane, where MeOS is likewise predicted to localize (Figure S7I). Furthermore, another enzyme, CoQ6, which performs essential priming hydroxylation prior to methoxylation of ubiquinol by CoQ3, is also upregulated in BQ cells (Figure S7C), presumably permitting MeOS to methoxylate HQs (Figure S7J). However, silencing *CoQ6* led to beetle lethality. Discovery of MeOS therefore provides a serendipitous genetic window into a mitochondrial route where 4-HB is modified by CoQ enzymes to make HQs (Figure 7I), closely paralleling ubiquinol synthesis

Figure 6. The BQ cell type

(A) Dmd domain structure with homology to Lac1 and 2.
(B) GC traces of gland secretions. Top: wild-type; bottom: *dmd* RNAi.
(C and D) *dmd* HCR (magenta; green, WGA).
(E and F) Single ion GC traces of HQ molecular ions 2-Me-1,4-HQ (*m/z* = 124) (E) and 2-MeO-3-Me-1,4-HQ (*m/z* = 154) (F) in wild-type (top) and *dmd*-RNAi beetles (bottom). Asterisks in (F) mark a non-target compound with *m/z* = 154.
(G) Unrooted ML tree of insect laccases with aleocharine-specific Dmd clade (LG+I+G4 model).
(H–J) Dmd activity assayed by GC-MS when purified protein is provided with HQs. GC traces show efficient BQ production in Dmd's presence (upper trace); lower trace is no Dmd control. HQs: 1,4-HQ (H); 2-Me-1,4-HQ (I); 2-MeO-3-Me-1,4-HQ (J).
(K) GC quantification of BQs produced by Dmd, Lac1, and Lac2 with different HQ substrates (n = 3 replicates/enzyme). Size of bar represent the mean, while error bars are standard error of the mean.
(L) 2-MeO-3-Me-1,4-HQ versus ubiquinol.
(M) Diagnostic ions for 4-HB-$^{13}C_6$ incorporation into 2-Me-1,4-BQ and 2-MeO-3-Me-1,4-BQ. Top trace: total ion current (TIC) wild-type chromatogram. Lower traces: diagnostic ions for BQs labeled with $^{13}C_6$ (*m/z* = 128 for 2-Me-1,4-BQ and *m/z* = 158 for 2-MeO-3-Me-1,4-BQ). Intensity of *m/z* = 128 and *m/z* = 158 ions was magnified 100× in control and 4-HB-$^{13}C_6$ fed beetles compared to TIC chromatogram.
(N) ML tree of CoQ3 enzymes (LG+I+G4 model) showing multiple copies in Aleocharinae (green boxed clade), including paralog MeOS.
(O) GC trace following *meos* RNAi shows specific loss of 2-MeO-3-Me-1,4-BQ.
(P) Model of BQ cell function: Tyr-derived 4-HB in mitochondria (MT) is modified to HQs via a ubiquinol-like pathway (Q). HQs are secreted into cell lumen for oxidation by Dmd. BQs travel to reservoir via duct.
(Q) Inferred HQ pathway (2-MeO-3-me-1,4-HQ as example); DC/OH, unidentified decarboxylase/hydroxylase.
See also Figures S6 and S7, Table S1, and Data S6.

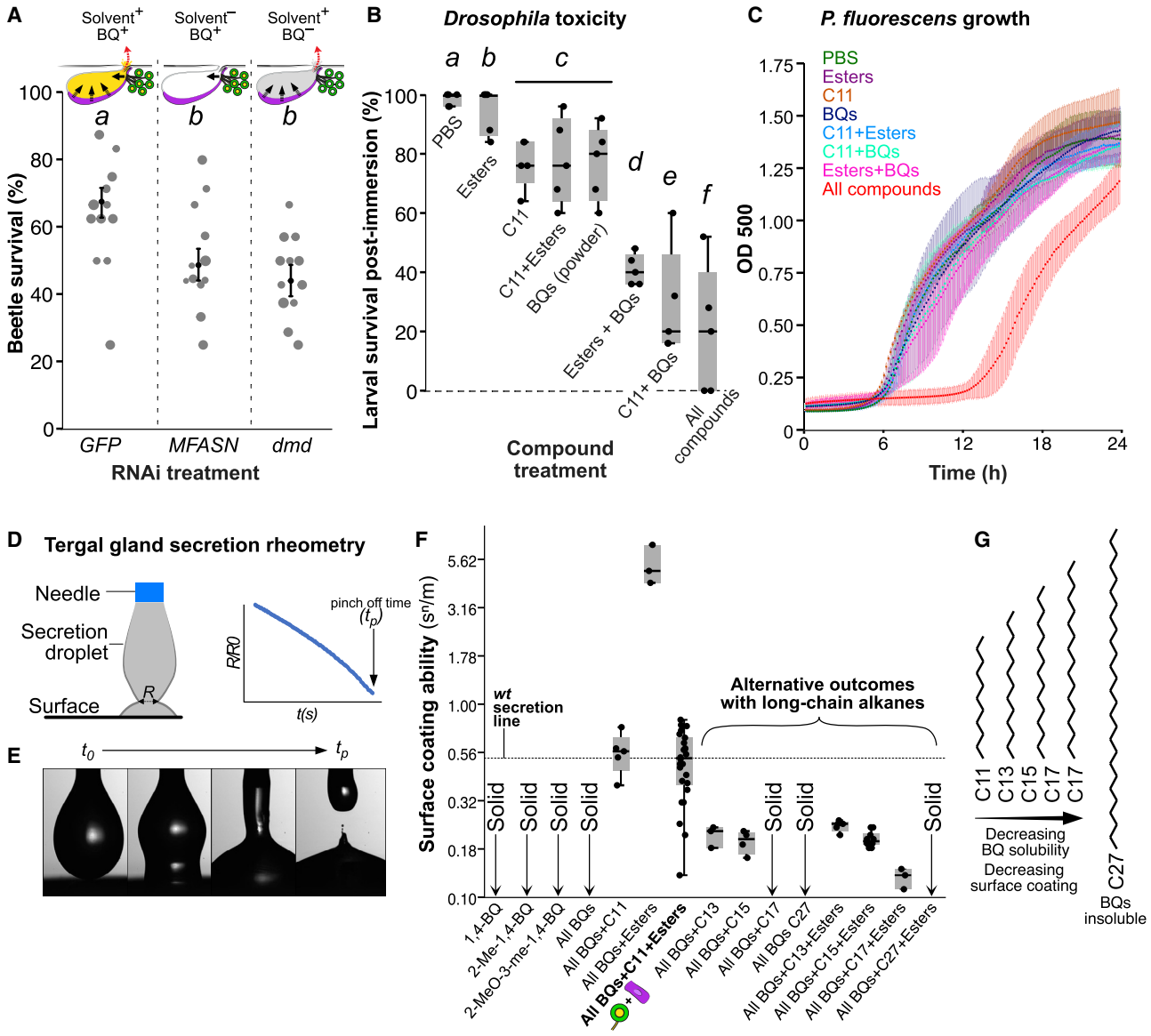


Figure 7. Synergism between solvent and BQ cells

(A) Survival with ants when BQ or solvent cells compromised (GLM, $X^2_{2,33} = 53.5$, $p < 0.001$). Dot size conveys number of armed beetles in arena included in statistical analysis. Black dots and bars are the estimates and SE of binomial model, respectively.

(B) Fly larva survival 1 h post-immersion in synthetic gland components (GLM, $X^2_{7,32} = 5842.8$, $p < 0.001$). Letters indicate significant differences. The middle bar of each boxplot represents the median, the box denotes the interquartile range, while the whiskers show the minimum and maximum.

(C) *P. fluorescens* growth curves over 24 h in presence of synthetic gland components ($ANOVA_{endpoint}$, $F_{7,96} = 6.9$, $p < 0.001$; Welch's $ANOVA_{lag\ time}$, $F_{7,96} = 134.9$, $p < 0.001$). Lines represent the mean for each growth curve; error bars denote the 95% confidence interval at each time point.

(D) Dripping-onto-substrate (DoS) rheometer: synthetic secretion is dispensed via nozzle onto aluminum substrate. Droplet diameter *R* is measured until droplet pinch off.

(E) Video sequence of DoS measurement.

(F) Surface coating ability ($K/\sigma \times 100$ in $s^n m^{-1}$) of mixtures. The middle bar of each boxplot represents the median, the box denotes the interquartile range, while the whiskers show the minimum and maximum.

(G) SCA and BQ solubility decrease with longer alkane chain length. See Table S1 for statistical analyses.

See also Data S5.

(Figure S7J). Fully delineating the pathway *in vivo* is currently impossible owing to essential systemic functions of other CoQ enzymes. However, we postulate their co-option as a parsimonious model and conjecture that *Dalotia*'s three BQs arise via differential processing by MeOS and other CoQ enzymes (Figures 6Q and S7J).

Coevolution of BQ and solvent cells

Why did cell type evolution follow the routes we have uncovered? The tergal gland secretion enhances beetle survival (Figure 1C), so we reasoned that the BQ and solvent cells together underlie the gland's adaptive value. To test this hypothesis, we performed a large-scale selection experiment, placing *Dalotia* in multiplexed arenas with predatory ants and quantifying relative survival of beetles that were either wild-type (GFP RNAi; 12 arenas, n = 120 beetles), *MFASN*-silenced to inhibit solvent production (12 arenas, n = 120 beetles), or *dmd*-silenced to inhibit BQ synthesis (12 arenas, n = 120 beetles). We allowed ants and beetles to interact for 48 h before assaying survival. In both *MFASN*- and *dmd*-silenced treatments, we observed a comparable, significant reduction in survival (64% for GFP RNAi, 50% for *MFASN* RNAi, $p < 0.001$, 45% for *dmd* RNAi; $p < 0.001$) (Figure 7A). Reduced survival appears to result from impaired chemical defense, because silencing *MFASN* or *dmd* without ant exposure did not significantly impact survival (Data S5A). These results demonstrate each cell type's adaptive value at the organismal level.

Because the BQ and solvent cells secrete into a common reservoir, their products may combine to make a bioactive secretion. We explored how the compounds interact to shape the secretion's physicochemical properties. Two parameters capture salient properties of defensive secretions: the surface coating ability (SCA)—a measure of wetting potential, and extensional viscosity (EV)—a substance's resistance to deformation when force is applied. The SCA and EV of arthropod secretions are species-specific and highly variable, likely arising as a balance of factors including safe containment in gland reservoirs, controlled exudation, efficient spread across target tissues, as well as final irritancy. We used a custom high-speed video rheometer to quantify SCA and EV of droplets of synthetic mixtures of different defensive compounds (Figures 7D and 7E). We found that SCAs of the BQ species alone or in combination are zero (Figure 7F), whereas their EVs are virtually infinite (Data S5B), because the three compounds remain in solid phase even when mixed. However, adding undecane dissolves the BQs, creating a moderately viscous liquid with high SCA (Figure 7F; Data S5B). Undecane thus acts as a solvent, unlocking the BQs. Equivalent SCA and EV are observed whether esters are present or not (Figure 7F). Undecane is therefore the primary solvent and the main determinant of the secretion's SCA and EV.

We examined the biological consequences of this chemical synergism by measuring toxicity against other organisms. Using viability of *Drosophila* larvae as a readout, we made synthetic secretions and measured survival rates 1 h after drastic, whole-body immersion. Dipping fly larvae for 1 s in a solution composed solely of undecane or esters, or both compound classes combined, led to only a small-to-modest survival reduction (Figure 7B). Due to the BQs being solids, we could not immerse fly larvae in them, but found that completely bathing larvae in BQ powder caused a comparably limited drop in survival (Figure 7B). However, when the BQs were combined with the undecane and esters at their natural ratios, mortality increased dramatically, reaching 80% (with 100% mortality in 2 of 5 replicates) (Figure 7B). High mortality was also obtained if undecane was substituted for a higher ester fraction, or vice versa, demonstrating the critical effect of dissolving the BQs (Figure 7B). These data indicate that the BQ and solvent cells are indeed engaged in biosynthetic synergism; only when their products are combined is a potent secretion with demonstrable adaptive value obtained (Figures 1C and 7A).

We studied the antimicrobial effects of the secretion, noting that *Dalotia* self-applies the secretion topically to its own body. To do so, we assayed how synthetic combinations impact growth of the Gram-negative bacterium, *Pseudomonas fluorescens*—a common soil species and potential insect pathogen (Pineda et al., 2010; Scales et al., 2014). When added to the *P. fluorescens* culture medium, no single compound inhibited growth (as measured by the OD500 of the culture at the endpoint of the assay) (Figure 7C; Table S1). Similarly, adding pairs of compound classes (undecane+esters, undecane+BQs, or esters+BQs) caused no effect (all pairwise tests $p \geq 0.18$) (Figure 7C; Table S1). Remarkably, however, all three compound classes together caused dramatic growth suppression (all pairwise tests $p \leq 0.007$) (Figure 7C). This suppressive effect was also manifested in a significantly prolonged lag-time prior to the exponential growth phase of the *P. fluorescens* culture (all pairwise tests $p < 0.001$) (Data S5C). Antimicrobial efficacy arising from combining these compounds is unexpected and striking; to our knowledge it has not been previously reported. The effect is again consistent with the adaptive utility of the tergal gland arising from cooperation between BQ and solvent cells.

Dalotia's tergal gland secretion is thus a multi-compound cocktail with emergent properties possessed by no single component. Although the BQs form the active defensive agent, the solvents provide a vehicle for the BQs while simultaneously conferring surface coating and tissue penetrating properties. We find two principal ways in which the solvent pathway may be specialized to meet these demands. First, the pathway favors high-level alkane and low-level ester synthesis. We noticed that silencing either *TG-CYP4G* or *TG-FAR* strongly increased levels of the C12 ester ("ester shift" in Figures 2H and 2I), consistent with TG-CYP4G and TG-FAR normally titrating most C12-CoA toward undecane, permitting TG-αEst to make only trace ethyl dodecanoate (Figures 1C and 2F). Accordingly, silencing *TG-CYP4G* or *TG-FAR* frees up C12-CoA for TG-αEst, resulting in elevated ethyl dodecanoate (Figures 2H and 2I). The alkane branch is thus specialized for C12-CoA, whereas TG-αEst can use both C10- and C12-CoA. In this way, the solvent pathway produces large amounts of undecane as both the primary solvent and determinant of EV and SCA. This alkane bias strongly impacts the secretion: if undecane is replaced with the esters at their natural ratios, substantially higher EV and exceptionally high SCA result (Figure 7F; Data S5B). Such a sticky, surface active secretion may be challenging to secrete and difficult to contain in the reservoir. Low abundance esters are nevertheless critical for the secretion's antimicrobial effect (Figure 7C) and may also promote penetration into cuticles (Dettner, 1991).

A second aspect is the use of undecane as opposed to a longer chain alkane as the solvent. Undecane stands in contrast to the very-long-chain CHCs produced in oenocytes. During solvent pathway evolution, the importance of selective recruitment of a reductase (FAR) and decarbonylase (CYP4G) but not an

elongase (ELO) is underscored by insolubility of BQs in even the shortest of *Dalotia*'s CHCs, heptacosane (C27) (Figures 7F and 7G). However, the lack of ELO recruitment was not the only critical feature. Most insect FASNs produce C14–C18 fatty acids, rather than C12. We find that the corresponding alkanes (C13, C15, and C17), although only moderately longer than undecane, create secretions with much lower SCAs when mixed with the BQs, even in the presence of esters (Figures 7F and 7G). In the case of heptadecane (C17), adding BQs without esters caused the secretion to freeze (Figure 7F), and BQ crystals were still evident when esters were included. Hence, synthesis of C12-CoA by MFASN is central to creating an alkane that can both solubilize the BQs and create a topical secretion (Figure 7G).

DISCUSSION

We have traced the evolution of new cellular functions comprising an organ novelty—the defensive tergal gland of rove beetles. First, we uncovered small molecule enzyme pathways that execute the biosynthetic functions of two taxon-restricted secretory cell types. Second, we presented evidence that these pathways—as well as a transcriptomic module expressed within one of the cell types—arose via repurposing components from more ancient cell types within the beetle. Third, we demonstrated the consequences of cell type evolution at the organ level, assigning adaptive value to the coordinated actions of the two cell types, as well as to specialized features of the solvent pathway that render the secretion effective in chemical defense. Our findings indicate that the functional evolution of animal cell types can be a constrained process, employing pre-existing pathway motifs and expression programs and hence potentially convergence-prone. Conversely, our study demonstrates that the route followed by cell-type evolution may also be a highly contingent process, conditional on other cell types within the organ and the fitness consequences of their collective output.

Evolution of new biosynthetic functions

Using scRNA-seq, we pinpointed specific cell types from which solvent pathway enzymes had been ancestrally sourced, enabling us to reconstruct the assembly of the pathway via co-option or duplication of oenocyte and fat body enzymes. We also determined that a transcriptome module that defines oenocytes and VFB, GEP9, has been re-employed in solvent cells. Through these changes, an ancestral region of abdominal cuticle evolved into specialized secretory cells comprising a reservoir. That the beetle's capacity to manufacture solvents is derived from its primary metabolism supports the view of genetic paths of least resistance to new animal chemistries—a notion stemming from pervasive convergence in the compounds that animals produce (Brückner and Parker, 2020). The BQ pathway likewise points to an underlying genomic blueprint; here, HQ synthesis, although not yet fully delineated, owes at least part of its existence to the ubiquinone pathway from which the *CoQ3* paralog *meos* originated via duplication. We predict the connection will turn out to be still more extensive, with ubiquinone and HQ pathways sharing a subset of enzymes (Figure S7J). Similarly, the role of the laccase, Dmd, parallels or may derive from the conserved function of Lac2 in cuticle tanning (Asano et al., 2019). The discovery of laccase involvement in BQ-based termite chemical defense (Bourguignon et al., 2016) hints at parallel use of laccases in BQ-producing arthropod clades. We propose that repurposing pathway motifs from ancient metabolic cell types—or entire expression programs that enable production of certain compounds—represent accessible routes to new biosynthetic functions in animal cell types. A consequence is that exploration of chemical space is constrained, leading to widespread convergence.

Cell-type evolution via transcriptomic hybridization

Together with recruitment of solvent pathway enzymes, we propose that acquisition of GEP9 was causal in transforming the intersegmental membrane between A6 and A7 into a glandular epithelium. The membrane is naturally invaginated between segments in rove beetles (Figure 5G); consequently, further growth created a reservoir into which the solvent cells could secrete directly (Figure 5H). The antiquity of the oenocytes and fat body, which originated at the base of the Insecta, combined with GEP9 being the sole program recovered in these cell types, indicates that the polarity of recruitment was from these ancient cell types into solvent cells rather than vice versa. We cannot, however, rule out that some newer GEP9 components may have evolved functions in solvent cells and found secondary, pleiotropic utility in oenocytes or VFB.

The solvent cells shed mechanistic light on the phenomenon of "cell type fusion" where a novel cell type appears to take on features of two ancestral cell types (Arendt et al., 2016; Oakley, 2017; Schlosser, 2018). Redeployment of a pre-existing transcriptional program in a new cellular context may be a common means for generating cellular- and organ-level novelties, and contrasts with duplication and divergence of sister cells within organs or repeating developmental fields such as body segments (Marioni and Arendt, 2017) or brain nuclei (Kebschull et al., 2020). Transcriptomic hybridization may be mechanistically facile if expression programs conferring cell identity are controlled by one or a few transcription factors (i.e., "terminal selectors") (Hobert, 2016). How GEP9 became expressed in solvent cells is unknown, but this abdominal location is a region of overlapping expression of the Abdominal A and Abdominal B Hox proteins, both of which are needed for solvent cell development (Parker et al., 2018). Conceivably, a terminal selector controlling GEP9 in oenocytes and VFB could have come under Hox control and been co-opted into cuticle cells, creating a GEP9/17 hybrid transcriptome. Future studies may uncover the regulatory basis of BQ and solvent cell identities.

Cell-type evolution shaped by cooperative interactions at the organ level

Demonstrating adaptive change at the molecular level depends on connecting such changes to phenotypic outcomes that differentially impact fitness (Barrett and Hoekstra, 2011). Our data connect the evolution of new cellular functions to a cooperative interaction between cell types that dictates whole organ performance and directly impacts animal survival. Hence, we propose that coevolution between the BQ and solvent cells has been driven, at least in part, by natural selection for organ-level

properties. Our findings underscore how metazoan cell type and organ evolution are intrinsically coupled, with changes at the cellular level being comprehensible—and conferring adaptive value—only when their impact at the collective, multicellular organ level is considered (Kishi and Parker, 2021).

Inferring the steps leading to cooperativity between the solvent and BQ cells is informed by two observations. First, the recalcitrance of the solid BQs makes a "solvents first" scenario likely, where alkanes or esters, or fatty acid-derived progenitors thereof, initially arose. Such compounds may have conferred modest chemical defense or acted as pheromones or lubricants for the flexible abdomen. This reasoning is supported by the second observation: that the earliest-diverging branch of the tergal gland-bearing Aleocharinae—the tribe Hypocyphtini—produce fatty acid-derivatives but lack BQs. We deduce that the prior evolution of a fatty acid-based secretion set the stage for the subsequent evolution of BQs, the potency of which could be unlocked by the former compounds acting as effective solvents. Features of the solvent pathway such as its alkane bias, the use of medium chain compounds and the presence of low-abundance esters, can be interpreted as specializations to harness the BQs' toxicity and manipulability (Figure 7F; Data S5). Hence, we suggest that the solvent cells in turn underwent adaptive modifications to make better use of the BQs, implying reciprocal coevolution between the two cell types. Across Aleocharinae, BQs are a constant feature of the secretion but the fatty acid derivatives vary extensively, including hydrocarbons, esters, and aldehydes of differing chain lengths and ratios (Steidle and Dettner, 1993). These differences are expected to strongly influence the secretion's physicochemical nature, further emphasizing reciprocal coevolution.

Cooperative behavior is a feature of biological systems at all scales of organization. The tergal gland, comprising the simplest case of only two cell types, presents a model for explaining the evolution of cooperativity at the organ level. According to this model, the solvent cells created a niche for the BQ cells, which enhanced the gland's adaptive value. The dependence of the BQs on the product of the solvent cells, and the reciprocal dependence of the solvent cells on the BQs to enhance their own adaptive value, meant that the two cell types became "locked in" as a unit evolving within constraints set by performance at the organ level. This hypothesized route toward cooperativity contrasts with models for cooperativity within protein complexes. Here, dependencies between subunits can arise via entrenchment of binding interactions that were ancestrally selectively neutral (Hochberg et al., 2020; Lukeš et al., 2011). In the tergal gland, we posit that interdependence has likewise been enforced, but through the addition of a new cell type that is functionally contingent on a preexisting cell type, which itself became obligately reliant on the second cell type to realize a relative selective advantage. Such a scenario could iterate through further cycles, informing how cooperative interactions between diverse cell types may arise generally within organs.

Limitations of study

The BQ cells' large diameter (~30–40 μm) and scarcity (~8–14 cells/animal) precluded their recovery in the 10x scRNA-seq cell type atlas. We were thus unable to determine their GEP composition via cNMF, and thus their transcriptomic relationship to other cell types remains unclear at this time. Future studies of early-branching aleocharines are needed to gauge the conservation of BQ and solvent pathway enzymes across phylogeny; such efforts may further illuminate the tergal gland's stepwise evolution.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Beetle husbandry
  - Microbial suppression
  - Biochemical tracer experiments and SIM mass spectrometry
  - Artificial disarming and survival biotest against ants
  - Double-stranded RNA preparation and RNAi knockdown in Dalotia
  - Extraction of gland compounds and gas chromatography/mass spectrometry
  - *In situ* hybridization chain reaction (HCR)
  - Fly toxicity biotest
  - Antimicrobial assay
  - Ant selection experiment
  - Bulk RNaseq and transcriptome assembly
  - Cell-type specific transcriptome sequencing (SMART-seq)
  - Cell-type specific transcriptome analysis
  - Preparation of single cell suspension for 10x scRNaseq
  - Single-cell RNA sequencing (scRNaseq) library preparation and sequencing
  - Analysis of scRNaseq data
  - Cell type annotation and statistical validation of cell classes
  - Consensus non-Negative Matrix Factorization
  - Genome assembly and gene prediction
  - RNA extraction and transcriptome assembly
  - Phylogenetic analysis
  - Protein Expression and *in vitro* enzymatic assays for laccase
  - Rheological measurements
- DATA ANALYSIS AND VISUALIZATION SOFTWARE

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2021.11.014.

ACKNOWLEDGMENTS

We thank Y. Kishi, T. Naragon, J. Wagner, M. Spero, C. VanDrisse, the Bioinformatics Resource Center, and the Single Cell Profiling and Engineering

Center (SPEC) in the Beckman Institute at Caltech for assistance with this work. We are grateful to M. Bronner, M. Dickinson and four anonymous reviewers for constructive feedback. A.B. is a Simons Fellow of the Life Sciences Research Foundation (LSRF). This work was supported by a Rita Allen Foundation Scholars Award, an Alfred P. Sloan Research Fellowship, a Shurl and Kay Curci Foundation grant, a Klingenstein-Simons Fellowship Award, and a National Science Foundation CAREER award (NSF 2047472 to J.P.).

AUTHOR CONTRIBUTIONS

Conceptualization, J.P. and A.B.; Methodology, J.P. and A.B.; Investigation, A.B., J.M.B., R.W.L., S.A.K., M.Y., and J.P.; Formal Analysis, A.B.; Data Curation, A.B.; Writing – Original Draft, J.P. and A.B.; Writing – Review & Editing, J.P., A.B., and S.A.K.; Supervision, J.P.; Project Administration, J.P.; Funding Acquisition, A.B. and J.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 23, 2021
Revised: September 29, 2021
Accepted: November 10, 2021
Published: December 9, 2021

REFERENCES

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Campbell, P.J., and Stratton, M.R. (2013). Deciphering signatures of mutational processes operative in human cancer. Cell Rep. *3*, 246–259.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. Austral Ecol. *26*, 32–46.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arakane, Y., Muthukrishnan, S., Beeman, R.W., Kanost, M.R., and Kramer, K.J. (2005). Laccase 2 is the phenoloxidase gene required for beetle cuticle tanning. Proc. Natl. Acad. Sci. USA *102*, 11337–11342.

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D., and Wagner, G.P. (2016). The origin and evolution of cell types. Nat. Rev. Genet. *17*, 744–757.

Arrese, E.L., and Soulages, J.L. (2010). Insect fat body: energy, metabolism, and regulation. Annu. Rev. Entomol. *55*, 207–225.

Asano, T., Seto, Y., Hashimoto, K., and Kurushima, H. (2019). Mini-review an insect-specific system for terrestrialization: Laccase-mediated cuticle formation. Insect Biochem. Mol. Biol. *108*, 61–70.

Babonis, L.S., Ryan, J.F., Enjolras, C., and Martindale, M.Q. (2019). Genomic analysis of the tryptome reveals molecular mechanisms of gland cell evolution. Evodevo *10*, 23.

Barrett, R.D.H., and Hoekstra, H.E. (2011). Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. *12*, 767–780.

Bates, D., and Maechler, M. (2010). Matrix: sparse and dense matrix classes and methods. R Package Version. https://cran.r-project.org/web/packages/Matrix/index.html.

Bates, D., Sarkar, D., Bates, M.D., and Matrix, L. (2007). The lme4 package. R Package Version *2*, 74.

Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Series B Stat Methodol *57*, 289–300.

Betz, O., Irmler, U., and Klimaszewski, J. (2018). Biology of Rove Beetles (Staphylinidae) (Springer).

Billeter, J.-C., Atallah, J., Krupp, J.J., Millar, J.G., and Levine, J.D. (2009). Specialized cells tag sexual and species identity in Drosophila melanogaster. Nature *461*, 987–991.

Blair, G.W.S., Hening, J.C., and Wagstaff, A. (1939). The Flow of Cream through Narrow Glass Tubes. J. Phys. Chem. *43*, 853–864.

Blomquist, G.J., and Ginzel, M.D. (2021). Chemical Ecology, Biochemistry, and Molecular Biology of Insect Hydrocarbons. Annu. Rev. Entomol. *66*, 45–60.

Blum, M. (1981). Chemical Defenses of Arthropods (Academic Press).

Bourguignon, T., Šobotník, J., Brabcová, J., Sillam-Dussès, D., Buček, A., Krasulová, J., Vytisková, B., Demianová, Z., Mareš, M., Roisin, Y., and Vogel, H. (2016). Molecular Mechanism of the Two-Component Suicidal Weapon of Neocapritermes taracua Old Workers. Mol. Biol. Evol. *33*, 809–819.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Breiman, L. (2001). Random Forests. Mach. Learn. *45*, 5–32.

Brückner, A., and Heethoff, M. (2017). A chemo-ecologists' practical guide to compositional data analysis. Chemoecology *27*, 33–46.

Brückner, A., and Parker, J. (2020). Molecular evolution of gland cell types and chemical interactions in animals. J. Exp. Biol. *223* (Pt, Suppl 1), jeb211938.

Brückner, A., Kaltenpoth, M., and Heethoff, M. (2020). De novo biosynthesis of simple aromatic compounds by an arthropod (*Archegozetes longisetosus*). Proc. Biol. Sci. *287*, 20201429.

Brückner, A., Barnett, A.A., Antoshechkin, I.A., and Kitchen, S.A. (2021). Molecular evolutionary trends and biosynthesis pathways in the Oribatida revealed by the genome of Archegozetes longisetosus. Biorxiv, 2020.12.10.420141.

Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. Nar. Genom. Bioinform. *3*, lqaa108.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Canty, A., and Ripley, B. (2017). boot: Bootstrap R (S-Plus) functions. R Package Version *1*, 3–20.

Carcamo-Noriega, E.N., Sathyamoorthi, S., Banerjee, S., Gnanamani, E., Mendoza-Trujillo, M., Mata-Espinosa, D., Hernández-Pando, R., Veytia-Bucheli, J.I., Possani, L.D., and Zare, R.N. (2019). 1,4-Benzoquinone antimicrobial agents against *Staphylococcus aureus* and *Mycobacterium tuberculosis* derived from scorpion venom. Proc. Natl. Acad. Sci. USA *116*, 12642–12647.

Carlson, M., and Pages, H. (2019). AnnotationForge: tools for building SQLite-based annotation data packages. R package version 1.36.0. https://bioconductor.org/packages/AnnotationForge.

Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation *in situ* hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753.

Clifford, R.J., Milillo, M., Prestwood, J., Quintero, R., Zurawski, D.V., Kwak, Y.I., Waterman, P.E., Lesho, E.P., and Mc Gann, P. (2012). Detection of bacterial 16S rRNA and identification of four clinically important bacteria by real-time PCR. PloS one *7*, e48558.

Crescente, J.M., Zavallo, D., Helguera, M., and Vanzetti, L.S. (2018). MITE Tracker: an accurate approach to identify miniature inverted-repeat transposable elements in large genomes. BMC Bioinformatics *19*, 348.

Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life (John Murray).

Dawkins, R. (1986). The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe Without Design (WW Norton & Company).

Dettner, K. (1984). Isopropylesters as Wetting Agents From the Defensive Secretion of the Rove Beetle Coprophilus striatulus F. (Coleoptera, Staphylinidae). Insect Biochem. *14*, 383–390.

Dettner, K. (1991). Solvent-dependent Variability of Effectiveness of Quinone-defensive Systems of Oxytelinae Beetles (Coleoptera: Staphylinidae). Entomol. Gen. *15*, 275–292.

Dettner, K. (1993a). Defensive secretions and exocrine glands in free-living staphylinid beetles—their bearing on phylogeny (Coleoptera: Staphylinidae). Biochem. Syst. Ecol. *21*, 143–162.

Dettner, K. (1993b). Dabbing and shooting of benzo- and naphthoquinone secretions: Defensive strategies of bark-inhabiting aleocharine (col.: Staphylinidae) and tenebrionid (col.: Tenebrionidae) beetle larvae. J. Chem. Ecol. *19*, 1337–1354.

Dettner, K., Schwinger, G., and Wunderle, P. (1985). Sticky secretion from two pairs of defensive glands of rove beetleDeleaster dichrous (Grav.) (Coleoptera: Staphylinidae). J. Chem. Ecol. *11*, 859–883.

Dinic, J., Jimenez, L.N., and Sharma, V. (2017). Pinch-off dynamics and dripping-onto-substrate (DoS) rheometry of complex fluids. Lab Chip *17*, 460–473.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doshi, P., Suryo, R., Yildirim, O.E., McKinley, G.H., and Basaran, O.A. (2003). Scaling in pinch-off of generalized Newtonian fluids. J Non-Newton Fluid *113*, 1–27.

Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., and Parsonage, H. (2019). Package 'data. table.' Extension of 'data (Frame).

Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191.

Dwivedi, U.N., Singh, P., Pandey, V.P., and Kumar, A. (2011). Structure–function relationship among bacterial, fungal and plant laccases. J. Mol. Catal. B Enzym. *68*, 117–128.

Eckart, C., and Young, G. (1936). The approximation of one matrix by another of lower rank. Psychometrika *1*, 211–218.

Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238.

Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

Erwin, D.H. (2015). Novelty and Innovation in the History of Life. Curr. Biol. *25*, R930–R940.

Erwin, D.H. (2021). A conceptual framework of evolutionary novelty and innovation. Biol. Rev. Camb. Philos. Soc. *96*, 1–15.

Feyereisen, R. (2020). Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Mol. Phylogenet. Evol. *143*, 106695.

Finet, C., Slavik, K., Pu, J., Carroll, S.B., and Chung, H. (2019). Birth-and-death evolution of the fatty acyl-CoA reductase (FAR) gene family and diversification of cuticular hydrocarbon synthesis in Drosophila. Genome Biol. Evol. *11*, 1541–1551.

Fox, J. (2003). Effect displays in R for generalised linear models. J. Stat. Softw. *8*, 1–27.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., and Graves, S. (2012). Package 'car' (R Foundation for Statistical Computing).

Francke, W., and Dettner, K. (2005). Chemical Signalling in Beetles. Top. Curr. Chem. *240*, 85–166.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

Griffith, O.W., and Wagner, G.P. (2017). The placenta as a model for understanding the origin and evolution of vertebrate organs. Nat. Ecol. Evol. *1*, 0072.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494–1512.

Hammer, O., Harper, D.A., and Ryan, P.D. (2001). PAST: Paleontological statistics software package for education and data analysis. Palaeontol. Electronica *4*, 9.

Haunerland, N.H., and Shirk, P.D. (1995). Regional and Functional Differentiation in the Insect Fact Body. Annu. Rev. Entomol. *40*, 121–145.

Hobert, O. (2016). Terminal Selectors of Neuronal Identity. Curr. Top Dev. Biol. *116*, 455–475.

Hochberg, G.K.A., Liu, Y., Marklund, E.G., Metzger, B.P.H., Laganowsky, A., and Thornton, J.W. (2020). A hydrophobic ratchet entrenches molecular complexes. Nature *588*, 503–508.

Holze, H., Schrader, L., and Buellesbach, J. (2021). Advances in deciphering the genetic basis of insect cuticular hydrocarbon biosynthesis and variation. Heredity *126*, 219–234.

Hu, Y., Linz, D.M., and Moczek, A.P. (2019). Beetle horns evolved from wing serial homologs. Science *366*, 1004–1007.

Ibarra, Y., and Blair, N.T. (2013). Benzoquinone reveals a cysteine-dependent desensitization mechanism of TRPA1. Mol. Pharmacol. *83*, 1120–1132.

Jorth, P., Spero, M.A., Livingston, J., and Newman, D.K. (2019). Quantitative Visualization of Gene Expression in Mucoid and Nonmucoid Pseudomonas aeruginosa Aggregates Reveals Localized Peak Expression of Alginate in the Hypoxic Zone. MBio *10*, e02622.

Kang, K., Pulver, S.R., Panzano, V.C., Chang, E.C., Griffith, L.C., Theobald, D.L., and Garrity, P.A. (2010). Analysis of Drosophila TRPA1 reveals an ancient origin for human chemical nociception. Nature *464*, 597–600.

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

Kawamukai, M. (2016). Biosynthesis of coenzyme Q in eukaryotes. Biosci. Biotechnol. Biochem. *80*, 23–33.

Kebschull, J.M., Richman, E.B., Ringach, N., Friedmann, D., Albarran, E., Kolluru, S.S., Jones, R.C., Allen, W.E., Wang, Y., Cho, S.W., et al. (2020). Cerebellar nuclei evolved by repeatedly duplicating a conserved cell-type set. Science *370*, eabd5059.

Keilwagen, J., Hartung, F., and Grau, J. (2019). GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. Methods Mol. Biol. *1962*, 161–177.

Kishi, Y., and Parker, J. (2021). Cell type innovation at the tips of the animal tree. Curr. Opin. Genet. Dev. *69*, 112–121.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187–1201.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B.E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., et al. (2016). *Jupyter Notebooks-a publishing format for reproducible computational workflows. In proceedings of the 20th International Conference on Electronic Publishing.* pp. 87-90.

Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. eLife *8*, e43803.

Koto, A., Motoyama, N., Tahara, H., McGregor, S., Moriyama, M., Okabe, T., Miura, M., and Keller, L. (2019). Oxytocin/vasopressin-like peptide inotocin regulates cuticular hydrocarbon synthesis and water balancing in ants. Proc. Natl. Acad. Sci. USA *116*, 5597–5606.

Kraaijeveld, K., Neleman, P., Mariën, J., Meijer, E. de, and Ellers, J. (2019). Genomic Resources for Goniozus legneri, Aleochara bilineata and Paykullia maculata, Representing Three Independent Origins of the Parasitoid Lifestyle in Insects. G3 (Bethesda) *9*, 987–991.

Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. F1000Res. *6*, 1287.

Lamb, T.D., Collin, S.P., and Pugh, E.N., Jr. (2007). Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. Nat. Rev. Neurosci. *8*, 960–976.

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., and Lam, T.-W. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods *102*, 3–11.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News *2/3*, 18–22.

Linz, D.M., and Tomoyasu, Y. (2018). Dual evolutionary origin of insect wings supported by an investigation of the abdominal wing serial homologs in *Tribolium*. Proc. Natl. Acad. Sci. USA *115*, E658–E667.

Litvaitis, M.K., and Rohde, K. (1999). A molecular test of platyhelminth phylogeny: inferences from partial 28S rDNA sequences. Invertebrate Biology *118*, 42–56.

Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F., and Gray, M.W. (2011). How a neutral evolutionary ratchet can build cellular complexity. IUBMB Life *63*, 528–537.

Lun, A.T.L., Riesenfeld, S., Andrews, T., Dao, T.P., Gomes, T., and Marioni, J.C.; 1st Human Cell Atlas Jamboree (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. *20*, 63.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience *1*, 18.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA *104* (*Suppl 1*), 8597–8604.

Makki, R., Cinnamon, E., and Gould, A.P. (2014). The development and functions of oenocytes. Annu. Rev. Entomol. *59*, 405–425.

Marioni, J.C., and Arendt, D. (2017). How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology. Annu. Rev. Cell Dev. Biol. *33*, 537–553.

Marshall, K.A., Liedtke, A.M., Todt, A.H., and Walker, T.W. (2017). Extensional rheometry with a handheld mobile device. Exp. Fluids *58*, 69.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. *17*, 10–12.

Maruyama, M., and Parker, J. (2017). Deep-Time Convergence in Rove Beetle Symbionts of Army Ants. Curr. Biol. *27*, 920–926.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv, arXiv:1802.03426.

Melsted, P., Ntranos, V., and Pachter, L. (2019). The barcode, UMI, set format and BUStools. Bioinformatics *35*, 4472–4473.

Melsted, P., Booeshaghi, A.S., Liu, L., Gao, F., Lu, L., Min, K.H.J., da Veiga Beltrame, E., Hjörleifsson, K.E., Gehring, J., and Pachter, L. (2021). Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat. Biotechnol. *39*, 813–818.

Meyer, D., Zeileis, A., Hornik, K., Gerber, F., Friendly, M., and Meyer, M.D. (2020). Package 'vcd.' Visualizing Categorical Data. R Package Version 1–4.

Morata, G., and Lawrence, P.A. (1975). Control of compartment development by the engrailed gene in Drosophila. Nature *255*, 614–617.

Morgan, M., Carlson, M., Tenenbaum, D., and Arora, S. (2019). Package AnnotationHub. https://bioconductor.org/packages/release/bioc/html/AnnotationHub.html.

Moses, L., and Pachter, L. (2021). BUSpaRse: kallisto | bustools R utilities, R Package Version 1.4.2.

Musser, J.M., Schippers, K.J., Nickel, M., Mizzon, G., Kohn, A.B., Pape, C., Hammel, J.U., Wolf, F., Liang, C., Hernández-Plaza, A., et al. (2021). Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. Science *374*, 717–723.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

Noh, M.Y., Muthukrishnan, S., Kramer, K.J., and Arakane, Y. (2016). Cuticle formation and pigmentation in beetles. Curr. Opin. Insect Sci. *17*, 1–9.

Oakley, T.H. (2017). Furcation and fusion: The phylogenetics of evolutionary novelty. Dev. Biol. *431*, 69–76.

Oakley, T.H., and Speiser, D.I. (2012). How Complexity Originates: The Evolution of Animal Eyes. Annu. Rev. Ecol. Evol. Syst. *46*, 1–24.

Parker, J. (2016). Myrmecophily in beetles (Coleoptera): evolutionary patterns and biological mechanisms. Myrmecol. News *22*, 65–108.

Parker, J. (2017). Staphylinids. Curr. Biol. *27*, R49–R51.

Parker, J., Eldredge, K.T., Thomas, I., Coleman, R., and Davis, S. (2018). Hox-logic of body plan innovations for social insect symbiosis in rove beetles. bioRxiv, 198945.

Petersen, M., Armisén, D., Gibbs, R.A., Hering, L., Khila, A., Mayer, G., Richards, S., Niehuis, O., and Misof, B. (2019). Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol. Biol. *19*, 11.

Philip, B.N., and Tomoyasu, Y. (2012). Gene knockdown analysis by double-stranded RNA injection. In *Molecular methods for evolutionary genetics* (Humana Press), pp. 471–497.

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc. Natl. Acad. Sci. USA *99*, 14002–14007.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690.

Pineda, A., Zheng, S.-J., van Loon, J.J.A., Pieterse, C.M.J., and Dicke, M. (2010). Helping plants to deal with insects: the role of beneficial soil-borne microbes. Trends Plant Sci. *15*, 507–514.

Pinheiro, J., and Bates, D. (2007). Linear and nonlinear mixed effects models. R Package Version *3*, 1–89.

Poon, W.W., Barkovich, R.J., Hsu, A.Y., Frankel, A., Lee, P.T., Shepherd, J.N., Myles, D.C., and Clarke, C.F. (1999). Yeast and rat Coq3 and Escherichia coli UbiG polypeptides catalyze both O-methyltransferase steps in coenzyme Q biosynthesis. J. Biol. Chem. *274*, 21665–21672.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE *5*, e9490.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. *35*, D61–D65.

Pryszcz, L.P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. *44*, e113.

Qiu, Y., Tittiger, C., Wicker-Thomas, C., Le Goff, G., Young, S., Wajnberg, E., Fricaux, T., Taquet, N., Blomquist, G.J., and Feyereisen, R. (2012). An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. Proc. Natl. Acad. Sci. USA *109*, 14858–14863.

Raff, R. (1996). The Shape of Life: Genes, Development, and the Evolution of Animal Form (University of Chicago Press).

Rambaut, A. (2012). FigTree v1.4.1. http://Tree.Biol.Ed.Ac.Uk/Software/Figtree/.

Ripley, B., Venables, B., Bates, D.M., Hornik, K., Gebhardt, A., Firth, D., and Ripley, M.B. (2013). Package 'mass.'. Cran R *538*, 113–120.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584.

Rosello, M., Sur, S., Barbet, B., and Rothstein, J.P. (2019). Dripping-onto-substrate capillary breakup extensional rheometry of low-viscosity printing inks. J. Non-Newton. Fluid *266*, 160–170.

Roussa, E. (2011). Channels and transporters in salivary glands. Cell Tissue Res. *343*, 263–287.

Sarkar, D. (2008). Lattice: Multivariate Data Visualization with R (Springer Science & Business Media).

Scales, B.S., Dickson, R.P., LiPuma, J.J., and Huffnagle, G.B. (2014). Microbiology, genomics, and clinical significance of the Pseudomonas fluorescens species complex, an unappreciated colonizer of humans. Clin. Microbiol. Rev. *27*, 927–948.

Schlosser, G. (2018). A Short History of Nearly Every Sense-The Evolutionary History of Vertebrate Sensory Cell Types. Integr. Comp. Biol. *58*, 301–316.

Schmidt-Rhaesa, A. (2007). The Evolution of Organ Systems (Oxford University Press).

Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. Nature *457*, 818–823.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

Smit, A.F. (2004). Repeat-Masker Open-3.0. http://www.Repeatmasker.Org.

Smit, A.F., and Hubley, R. (2008). RepeatModeler Open-1.0. https://www.scienceopen.com/document?vid=c1050e3d-8dc1-48ba-81c8-b372af25be50.

Smith, S. (1994). The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. FASEB J. *8*, 1248–1259.

Smith, S., Witkowski, A., and Joshi, A.K. (2003). Structural and functional organization of the animal fatty acid synthase. Prog. Lipid Res. *42*, 289–317.

Sousa, P.C., Vega, E.J., Sousa, R.G., Montanero, J.M., and Alves, M.A. (2017). Measurement of relaxation times in extensional flow of weakly viscoelastic polymer solutions. Rheol. Acta *56*, 11–20.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19* (*Suppl 2*), ii215–ii225.

Stanley-Samuelson, D.W., Jurenka, R.A., Cripps, C., Blomquist, G.J., and de Renobales, M. (1988). Fatty acids in insects: Composition, metabolism, and biological significance. Arch. Insect Biochem. *9*, 1–33.

Stefely, J.A., and Pagliarini, D.J. (2017). Biochemistry of Mitochondrial Coenzyme Q Biosynthesis. Trends Biochem. Sci. *42*, 824–843.

Steidle, J.L.M., and Dettner, K. (1993). Chemistry and morphology of the tergal gland of freeliving adult Aleocharinae (Coleoptera: Staphylinidae) and its phylogenetic significance. Syst. Entomol. *18*, 149–168.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Surm, J.M., and Moran, Y. (2021). Insights into how development and life-history dynamics shape the evolution of venom. Evodevo *12*, 1.

Suzuki, Y., Takiya, S., Suzuki, T., Hui, C., Matsuno, K., Fukuta, M., Nagata, T., and Ueno, K. (1990). Developmental regulation of silk gene expression in Bombyx mori. In Molecular Insect Science, H.H. Hagedorn, J.G. Hildebrand, M.G. Kidwell, and J.H. Law, eds. (SpringerLink), pp. 83–89.

Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., and Borodovsky, M. (2008). Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. *18*, 1979–1990.

Tomoyasu, Y., and Denell, R.E. (2004). Larval RNAi in Tribolium (Coleoptera) for analyzing adult development. Dev. Genes Evol. *214*, 575–578.

Torres, J.P., and Schmidt, E.W. (2019). The biosynthetic diversity of the animal world. J. Biol. Chem. *294*, 17684–17692.

Torres, J.P., Lin, Z., Winter, J.M., Krug, P.J., and Schmidt, E.W. (2020). Animal biosynthesis of complex polyketides in a photosynthetic partnership. Nat. Commun. *11*, 2882.

VanDrisse, C.M., and Escalante-Semerena, J.C. (2016). New high-cloning-efficiency vectors for complementation studies and recombinant protein overproduction in Escherichia coli and Salmonella enterica. Plasmid *86*, 1–6.

Van Rossum, G., & Drake, F. L. (2000). *Python reference manual*. (iUniverse).

Vucic-Pestic, O., Birkhofer, K., Rall, B.C., Scheu, S., and Brose, U. (2010). Habitat structure and prey aggregation determine the functional response in a soil predator–prey interaction. Pedobiologia (Jena) *53*, 307–312.

Wagner, G.P., and Lynch, V.J. (2010). Evolutionary novelties. Curr. Biol. *20*, R48–R52.

Wagner, J.M., Naragon, T.H., and Brückner, A. (2020). Benzoquinones in the defensive secretion of a bug (Pamillia behrensii): a common chemical trait retrieved in the Heteroptera. bioRxiv. https://doi.org/10.1101/2020.12.11.421891.

Wakil, S.J., Stoops, J.K., and Joshi, V.C. (1983). Fatty acid synthesis and its regulation. Annu. Rev. Biochem. *52*, 537–579.

Wang, Y., and Hekimi, S. (2019). The Complexity of Making Ubiquinone. Trends Endocrinol. Metab. *30*, 929–943.

Way, G.P., and Greene, C.S. (2019). Discovering Pathway and Cell Type Signatures in Transcriptomic Compendia with Machine Learning. Annu. Rev. Biomed. Data Sci. *2*, 1–17.

West-Eberhard, M.J. (2003). Developmental Plasticity and Evolution (Oxford University Press).

Wickham, H. (2011). ggplot2. Wiley Interdiscip. Rev. Comput. Stat. *3*, 180–185.

Wickham, H., and Ruiz, E. (2019). dbplyr: A "dplyr" Back End for Databases. R package Version 1.4. 0.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., and Hester, J. (2019). Welcome to the Tidyverse. J. Open Source Softw. *4*, 1686.

Wilke, C. (2017). Package 'cowplot': Streamlined Plot Theme and Plot Annotations for 'ggplot2.' CRAN.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Dalotia coriaria	Applied Bionomics (Canada)	https://www.appliedbio-nomics.com/products/dalotia/
Drosophila melanogaster	Dickinson Laboratory (Caltech)	N/A
Liometopum occidentale	This study (field collection)	N/A
Bacterial and viral strains		
Pseudomonas fluorescens wild type (strain WCS365)	Jorth et al., 2019	N/A
BL21 competent *E. coli*	provided by C. VanDrisse, Caltech	N/A
Chemicals, peptides, and recombinant proteins		
amoxicillin	Sigma-Aldrich	#Cat A8523-5G
streptomycin sulfate salt	Sigma-Aldrich	#Cat S6501-25G
tetracycline	Sigma-Aldrich	#Cat 87128-25G
sodium dodecyl sulfate ACS reagent, ≥99.0%	Sigma-Aldrich	#Cat 436143-100G
13C6-D-glucose ≥99 atom % 13C, ≥99% (CP)	Sigma-Aldrich	#Cat 389374-1G
13C6-tyrosine >99 atom % 13C, 99% (CP)	Sigma-Aldrich	#Cat 489794-100MG
D3-methionine (methyl-D3, 98%)	Cambridge Isotope Laboratories	#Cat DLM-431-1
D23-dodecanoid acid ≥98 atom % D, ≥98% (CP)	Sigma-Aldrich	#Cat 451401-25MG
13C6-4-hydroxybenzoic acid (RING-13C6, 99%)	Cambridge Isotope Laboratories	#Cat CLM-273-0.5
hexane ReagentPlus, ≥99%	Sigma-Aldrich	#Cat 139386-500ML
diethyl pyrocarbonate (DEPC)	Sigma-Aldrich	#Cat 40718-25ML
1,4-benzoquinone, Certified Reference Material	Sigma-Aldrich	#Cat PHR1028-1G
2-methyl-1,4-benzoquinone 98%	Sigma-Aldrich	#Cat 211311-5G
2-methoxy-3-methyl-1,4-benzoquinone	This study; see below	N/A
1,4-hydroquinone ReagentPlus, 99%	Sigma-Aldrich	#Cat H17902-100G
2-methyl-1,4- hydroquinone 99%	Sigma-Aldrich	#Cat 112968-250G
2-methoxy-3-methyl-1,4- hydroquinone	Sigma-Aldrich	#Cat L300217-1G
n-undecane, reference substance for GC	Sigma-Aldrich	#Cat 1097940005
isopropyl decanoate AldrichCPR	Sigma-Aldrich	#Cat CDS000590
ethyl decanoate ReagentPlus, ≥99%	Sigma-Aldrich	#Cat 148970-100ML
ethyl dodecanoate ≥98.0% (GC)	Sigma-Aldrich	#Cat 61630-100ML
octadecane analytical standard	Sigma-Aldrich	#Cat 74691-5G
16% paraformaldehyde aqueous solution	Electron Microscopy Sciences	#Cat 15710
Trizol	Thermo Fisher	#Cat 15596026
Schneider's *Drosophila* medium	Thermo Fisher	#Cat 21720024
FBS - fetal bovine serum	Thermo Fisher	#Cat A3160501
Corning 0.25% Trypsin 0.1% EDTA in HBSS	Fisher Scientific	#Cat MT25053CI
Gibco Defined Trypsin Inhibitor	Sigma-Aldrich	#Cat R007100
Dulbecco's Phosphate Buffered Saline	Sigma-Aldrich	#Cat D8537-100ML
Bovine Serum Albumin solution	Sigma-Aldrich	#Cat B8667-1.25ML
Earle's Balanced Salt Solution	Sigma-Aldrich	#Cat E7510-100ML
Isopropyl-β-D-thiogalactoside (IPTG)	Sigma-Aldrich	#Cat 10724815001
Ni-NTA resin	Sigma-Aldrich	#Cat 70666-4
imidazole ReagentPlus, 99%	Sigma-Aldrich	#Cat I202-100G
DNase grade II, from bovine pancreas	Sigma-Aldrich	#Cat 10104159001

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
2-Morpholinoethanesulfonic acid monohydrate EMPROVE EXPERT	Sigma-Aldrich	#Cat 1370740250
SnakeSkin Dialysis Tubing, 10K MWCO, 16 mm	Thermo Fisher	#Cat 68100
Copper (II) sulfate ReagentPlus, ≥99%	Sigma-Aldrich	#Cat C1297-100G
ABTS (2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid))	Sigma-Aldrich	#Cat 10102946001
urea, BioReagent, for molecular biology,	Sigma-Aldrich	#Cat U5378-100G
sodium sulfate anhydrate - Na2SO4	Sigma-Aldrich	#Cat 238597-500G
dichloromethane, ≥99.8%, with amylene stabilizer	Sigma-Aldrich	#Cat 34856-1L
sodium periodate ACS reagent, ≥99.8%	Sigma-Aldrich	#Cat 311448-100G
tridecane analytical standard	Sigma-Aldrich	#Cat 91490-5ML
pentadecane analytical standard	Sigma-Aldrich	#Cat 442700
heptadecane analytical standard	Sigma-Aldrich	#Cat 51578-5ML
heptacosane analytical standard	Sigma-Aldrich	#Cat 51559-250MG
chloroform, ≥99.8%, with amylene stabilizer	Sigma-Aldrich	#Cat 372978-1L
Hoechst 33342 Solution (20 mM)	Thermo Fisher	#Cat 62249
Alexa-488-Wheat Germ Agglutinin Conjugate	Thermo Fisher	#Cat W11261
Alexa-647-Wheat Germ Agglutinin Conjugate	Thermo Fisher	#Cat W32466
BODIPY 493/503	Thermo Fisher	#Cat D3922
ProLong Gold Antifade Mountant	Thermo Fisher	#Cat P36934
RNAlater™ Stabilization Solution	Thermo Fisher	#Cat AM7020
Critical commercial assays		
Chromium Single Cell 3′ GEM, Library & Gel Bead Kit v3	10x Genomics	#Cat 1000075
Chromium Chip B Single Cell Kit v3	10x Genomics	#Cat 1000073
Chromium i7 Multiplex Kit	10x Genomics	#Cat 120262
NEBNext Single Cell/Low Input RNA Library Prep Kit for Illumina	New England Biolabs	#Cat E6420L
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1)	New England Biolabs	#Cat E7600S
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 2)	New England Biolabs	#Cat E7780S
Zymo Research Quick-DNA Miniprep Plus Kit	Zymo Research Corporation	#Cat D4069
Luna Universal qPCR Master Mix	New England Biolabs	#Cat M3003L
Qubit dsDNA HS Assay Kit	Thermo Fisher	#Cat Q32854
TOPO TA Cloning Kit	Thermo Fisher	#Cat 450641
MEGAscript T7 Transcription Kit	Thermo Fisher	#Cat AMB13345
In situ hybridization chain reaction (HCR) v3 kit for O-CYP4G-B1	Molecular Instruments	3752/D845
In situ HCR v3 kit for dmd-B1	Molecular Instruments	3752/D837
In situ HCR v3 kit for MFASN-B1	Molecular Instruments	3768/D873
In situ HCR v3 kit for TG-FAR-B1	Molecular Instruments	3752/D847
In situ HCR v3 kit for TG-CYP4G I-B1	Molecular Instruments	3834/E003 and 3752/D853
In situ HCR v3 kit for TG-CYP4G II-B4	Molecular Instruments	4096/E272 and 4096/E274
In situ HCR v3 kit for TG-α-EST-B1	Molecular Instruments	3834/E004 and 3752/D851
In situ HCR v3 kit for Tcas_MFASN-B1	Molecular Instruments	4096/E270
Illumina TruSeqTM mRNA	Illumina	#Cat 20020594
Agilent Bioanalyzer High Sensitivity DNA assay (Analysis Kit and DNA Chips)	Agilent	#Cat 5067-4626 / 5067-4627
LIVE/DEAD Viability/Cytotoxicity Kit	Thermo Fisher	#Cat L3224

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Blood & Cell Culture DNA Mini Kit (25)	Qiagen	#Cat 13323
Illumina TruSeqTM DNA kit v2	Illumina	#Cat FC-121-2001
NEBNext Ultra II FS DNA Library Prep Kit for Illumina	New England Biolabs	#Cat E7805L
Oligonucleotides		
Univ16SRT-F: 5′-ACTCCTACGGGAGGCAGCAGT-3′	Clifford et al., 2012	N/A
Univ16SRT-R: 5′-TATTACCGCGGCTGCTGGC-3′	Clifford et al., 2012	N/A
D3A _F: 5′-GACCCGTCTTGAAACACGGA-3′	Litvaitis and Rohde, 1999	N/A
D3B_R: 5′-TCGGAAGGAACCAGCTACTA-3′	Litvaitis and Rohde, 1999	N/A
HCR v3 Amplifier Hairpin System Alexa546-B1	Molecular Instruments	N/A
HCR v3 Amplifier Hairpin System Alexa488-B4	Molecular Instruments	N/A
pTEV16 protein expression vector	VanDrisse and Escalante-Semerena, 2016	provided by C. VanDrisse, Caltech
RNAi target primers	This study	Sequence (5′ – 3′)
O-CYP4G _F1	This study; Integrated DNA Technologies, Inc	CACTCCCTGTCGGAACCTTGGA
O-CYP4G _R1	This study; Integrated DNA Technologies, Inc	TTGCGACATCCTCCACAGACGT
O-CYP4G _F2	This study; Integrated DNA Technologies, Inc	ACGTCTGTGGAGGATGTCGCAA
O-CYP4G _R2	This study; Integrated DNA Technologies, Inc	ATCCAAAATCCCCGGACCCGAT
TG-CYP4G_F3	This study; Integrated DNA Technologies, Inc	GAACATCCTTGCGTTCCTCAAC
TG-CYP4G _R3	This study; Integrated DNA Technologies, Inc	ATCTGACAGAGAATGCACTTTTGC
TG-CYP4G _F4	This study; Integrated DNA Technologies, Inc	ATGGCCGTTGTAAAGCTTTCGG
TG-CYP4G _R4	This study; Integrated DNA Technologies, Inc	GGCGGTGGTATCATGACCTTCA
TG-CYP4G _F5	This study; Integrated DNA Technologies, Inc	CCAGATGTAATGAATGGCTAGGGC
TG-CYP4G _R5	This study; Integrated DNA Technologies, Inc	CCTGTGCGGAAGGCTAATATTG
ELO_708_F1	This study; Integrated DNA Technologies, Inc	TCCGCTCATGGGTAATTGCTCG
ELO_708_R1	This study; Integrated DNA Technologies, Inc	TTTTGCCATCGAACATCACCGCA
ELO_708_F2	This study; Integrated DNA Technologies, Inc	AAATGGATGGCGAATAGACCCG
ELO_708_R2	This study; Integrated DNA Technologies, Inc	GCAGGTTGTGCGGGTTTTGTTT
ELO_804_F3	This study; Integrated DNA Technologies, Inc	GGCTGGGAGTGACAAAACTCGG
ELO_804_R3	This study; Integrated DNA Technologies, Inc	TCCGCAGATGAGGTCCTAGTGC
ELO_804_F4	This study; Integrated DNA Technologies, Inc	TTGCTGGTGGCCATGCAACTTT
ELO_804_R4	This study; Integrated DNA Technologies, Inc	TCTTGGACATGCCATCCGACGA
Dmd_F1	This study; Integrated DNA Technologies, Inc	TCATCCATAGCTCGTACCGTTG
Dmd_R1	This study; Integrated DNA Technologies, Inc	TTAGTTTCTGTGACGTCCCTGG
Dmd_F2	This study; Integrated DNA Technologies, Inc	ATGTGCAGGCAAATCTTGATCG
Dmd_R2	This study; Integrated DNA Technologies, Inc	TGGCACGTGTTTGTTCATACAC
Dmd_F3	This study; Integrated DNA Technologies, Inc	CCAGGGACGTCACAGAAACTAA
Dmd_R3	This study; Integrated DNA Technologies, Inc	ATTGGACTATAAATGGCCGCCT
Lac_2 _long_F1	This study; Integrated DNA Technologies, Inc	CCTGGAGGTGACCACGTCTTGA
Lac_2_long_R1	This study; Integrated DNA Technologies, Inc	CTCCTATAGGCGTTGGCGGAGT
Lac_2_ short_F2	This study; Integrated DNA Technologies, Inc	GCCAGTTCCGTGATCCCAACAC
Lac_2_ short_R2	This study; Integrated DNA Technologies, Inc	GATCTTGACATCTGGACGCGCC
MeOS_F1	This study; Integrated DNA Technologies, Inc	TCCTGGTTTGGACAGTCTGCGA
MeOS_R1	This study; Integrated DNA Technologies, Inc	CGTCGAGCGTCGAGCACTTATG
TG-α-EST_F1	This study; Integrated DNA Technologies, Inc	CGGTTTCCTCAACCTCAAAGACAC
TG-α-EST_R1	This study; Integrated DNA Technologies, Inc	GTCTGCGTCGTTCTTTGGAGTG
TG-α-EST_F2	This study; Integrated DNA Technologies, Inc	TGCCCAAACACAACTTGCAACT
TG-α-EST_R2	This study; Integrated DNA Technologies, Inc	CTCCACAACCACCGAAACCATG
O-FASN_708_F1	This study; Integrated DNA Technologies, Inc	AGTACGCCAGGTTTCTCCGTGT
O-FASN_708_R1	This study; Integrated DNA Technologies, Inc	AACAGCGCAACCAGTTGTACCC

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
O-FASN_708_F2	This study; Integrated DNA Technologies, Inc	TGCCAATGGTCCTCCCCATGAA
O-FASN_708_R2	This study; Integrated DNA Technologies, Inc	TCAGCTCTGAGGACAGCTGCTC
O-FASN_1114_F3	This study; Integrated DNA Technologies, Inc	GTCTAGCATCACCAGCACGCTG
O-FASN_1114_R3	This study; Integrated DNA Technologies, Inc	AGTTCCCGATGCTCCTGAACGT
O-FASN_1114_F4	This study; Integrated DNA Technologies, Inc	GGGAGCTTCTTCTCCTCCGCTT
O-FASN_1114_R4	This study; Integrated DNA Technologies, Inc	AACGCGGGTCAAACCAACTACG
O-FASN_1220_F5	This study; Integrated DNA Technologies, Inc	GGTGGCACGTTGCAACAAAAGG
O-FASN_1220_R5	This study; Integrated DNA Technologies, Inc	CTGTAGTCCGCATCAACGCCAG
O-FASN_1220_F6	This study; Integrated DNA Technologies, Inc	GATCGACCACCCTGAGATCCCC
O-FASN_1220_R6	This study; Integrated DNA Technologies, Inc	TCCGAGCGGGTATAACCAGTCG
O-FAR_934_F1	This study; Integrated DNA Technologies, Inc	CAGGCGGAACTGGTTTCCTTGG
O-FAR_934_R1	This study; Integrated DNA Technologies, Inc	GTACAATGGCGACCGGAAGTGG
O-FAR_934_F2	This study; Integrated DNA Technologies, Inc	GTGTAACATTCAGCGTGGGCGT
O-FAR_934_R2	This study; Integrated DNA Technologies, Inc	ATGCGGCATCTTCCATTCTCGC
O-FAR_1142_F3	This study; Integrated DNA Technologies, Inc	ACGCATGCCACAATAACCACCC
O-FAR_1142_R3	This study; Integrated DNA Technologies, Inc	ACGCGACGAACATCCAATGGAC
O-FAR_1142_F4	This study; Integrated DNA Technologies, Inc	TGCCATAATCCCCAACAAGTTTGGC
O-FAR_1142_R4	This study; Integrated DNA Technologies, Inc	CGGATTTGTTGGTAGGGCGCTT
MFASN_F1	This study; Integrated DNA Technologies, Inc	AGCAATTTCTTGGCACCCGCTT
MFASN_R1	This study; Integrated DNA Technologies, Inc	TCTAGCAACTCTGGCTTGGGCA
TG-FAR_F1	This study; Integrated DNA Technologies, Inc	GCACTGCCAGGTTCAACGTC
TG-FAR_R1	This study; Integrated DNA Technologies, Inc	CCAACTGACCCTAGAGAATTACTC
TG-FAR_F2	This study; Integrated DNA Technologies, Inc	CGGAAGAGTTTGCTTGTCACTG
TG-FAR_R2	This study; Integrated DNA Technologies, Inc	ACGCACTTGACTATATGCTGTACT
Software and Algorithms		
R v.3.6.0	R Core Team, 2018	https://cran.r-project.org/bin/windows/base/old/3.6.0/
Past: Paleontological Statistics Software Package for Education and Data Analysis 4.04	Hammer et al., 2001	https://www.nhm.uio.no/english/research/infrastructure/past/downloads/past4.04.zip
Phyton v3	Van Rossum and Drake, 2000	https://www.python.org/download/releases/3.0/
TransDecoder	Haas et al., 2013	https://github.com/TransDecoder/TransDecoder/wiki
cd-hit	Fu et al., 2012	http://weizhongli-lab.org/cd-hit/
Orthofinder v2	Emms and Kelly, 2015, 2019	https://github.com/davidemms/OrthoFinder
Diamond	Buchfink et al., 2015	https://github.com/bbuchfink/diamond
MCL clustering	Enright et al., 2002	https://metacpan.org/pod/Bio::Tools::Run::TribeMCL
MAFFT	Katoh and Standley, 2013	https://mafft.cbrc.jp/alignment/software/
FastTree 2	Price et al., 2010	http://www.microbesonline.org/fasttree/
IQ-TREE v1.6.12	Nguyen et al., 2015	http://www.iqtree.org/
FigTree v1.4.4	Rambaut, 2012	http://tree.bio.ed.ac.uk/software/figtree/
REViGO tool	http://revigo.irb.hr	http://revigo.irb.hr/
MonaGO tool	https://monago.erc.monash.edu/.s17	https://monago.erc.monash.edu/
cNMF	Kotliar et al., 2019	https://github.com/dylkot/cNMF

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Seurat v3.1	Butler et al., 2018; Stuart et al., 2019	https://cran.r-project.org/web/packages/Seurat/index.html
BUStools (v.0.39.3)	Melsted et al., 2021	https://bustools.github.io/about
sleuth 0.30.0	Pimentel et al., 2017	https://github.com/pachterlab/sleuth
kallisto 0.46.0	Bray et al., 2016	https://pachterlab.github.io/kallisto/
BUSCO v4.1.1	Simão et al., 2015	https://gitlab.com/ezlab/busco
BLAST	Altschul et al., 1990	https://blast.ncbi.nlm.nih.gov/Blast.cgi
Trinity v2.8.4	Grabherr et al., 2011	https://github.com/trinityrnaseq/trinityrnaseq
STAR 2.7.0a	Dobin et al., 2013	https://github.com/alexdobin/STAR
JupyterLab	Kluyver et al., 2016	https://jupyter.org/
AnnotationForge 1.34.0	Carlson and Pages, 2019	https://bioconductor.org/packages/release/bioc/html/AnnotationForge.html
AnnotationHub 3.0.0	Morgan et al., 2019	https://bioconductor.org/packages/release/bioc/html/AnnotationHub.html
Biomart 2.48	Durinck et al., 2009	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
boot	Canty and Ripley, 2017	https://cran.r-project.org/web/packages/boot/boot.pdf
BUSpaRse 1.0	Moses and Pachter, 2021	https://bioconductor.org/packages/release/bioc/html/BUSpaRse.html
Car 3.0-10	Fox et al., 2012	https://cran.r-project.org/web/packages/car/index.html
clusterProfiler 2.8	Yu et al., 2012	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Cowplot 1.1.1	Wilke, 2017	https://cran.r-project.org/web/packages/cowplot/index.html
data.table 1.12.3	Dowle et al., 2019	https://cran.r-project.org/web/packages/data.table/data.table.pdf
Dbplyr 1.4.3	Wickham and Ruiz, 2019	https://dbplyr.tidyverse.org/
DropletUtils 1.8.0	Lun et al., 2019	https://bioconductor.org/packages/release/bioc/html/DropletUtils.html
Effects 4.2-0	Fox, 2003	https://cran.r-project.org/web/packages/effects/effects.pdf
ggplot2 3.3.0	Wickham, 2011	https://ggplot2.tidyverse.org/
lattice 0.20-44	Sarkar, 2008	https://cran.r-project.org/web/packages/lattice/lattice.pdf
lme4 1.1	Bates et al., 2007	https://cran.r-project.org/web/packages/lme4/lme4.pdf
MASS 7.3	Ripley et al., 2013	https://cran.r-project.org/web/packages/MASS/index.html
Matrix 1.2	Bates and Maechler, 2010	https://cran.r-project.org/web/packages/Matrix/Matrix.pdf
nlme 3.1-108	Pinheiro and Bates, 2007	https://cran.r-project.org/web/packages/nlme/nlme.pdf
randomForest 4.6-16	Liaw and Wiener, 2002	https://cran.r-project.org/web/packages/randomForest/randomForest.pdf
tidyverse 1.2.1	Wickham et al., 2019	https://www.tidyverse.org/
vcd 1.3-2	Meyer et al., 2020	https://cran.r-project.org/web/packages/vcd/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FastQC v0.11.8	Andrews, 2010	https://github.com/s-andrews/FastQC/blob/master/RELEASE_NOTES.txt
cutadapt v1.18	Martin, 2011	https://kbase.us/applist/apps/kb_cutadapt/remove_adapters/release
MEGAHIT v1.1.3	Li et al., 2016	https://github.com/voutcn/megahit
Redundans v0.14a	Pryszcz and Gabaldón, 2016	https://github.com/lpryszcz/redundans
GapCloser v1.12	Luo et al., 2012	https://github.com/BGI-Qingdao/stLFR_GapCloser
RepeatModeler v 1.0.11	Smit and Hubley, 2008	https://www.repeatmasker.org/RepeatModeler/
MITE tracker	Crescente et al., 2018	https://github.com/INTABiotechMJ/MITE-Tracker
vsearch v 2.7.1	Rognes et al., 2016	https://bioweb.pasteur.fr/packages/pack@vsearch@2.7.1/
RepeatMasker v 4.07	Smit, 2004	https://www.repeatmasker.org/
GeneMark-ES v.4.33	Ter-Hovhannisyan et al., 2008	http://exon.gatech.edu/GeneMark/
Augustus v3.2.3	Stanke and Waack, 2003	http://augustus.gobics.de/
BRAKER v2.1.2	Brůna et al., 2021	https://github.com/Gaius-Augustus/BRAKER
PASA v2.3.3	Haas et al., 2008	https://github.com/PASApipeline/PASApipeline/releases
GeMoMA v1.6.1	Keilwagen et al., 2019	http://www.jstacs.de/index.php/GeMoMa
EVidenceModeler v.1.1.1	Haas et al., 2008	https://evidencemodeler.github.io/
Data analysis scripts	This study	CaltechData: https://doi.org/10.22002/D1.1918
Deposited data		
processed scRNAseq 10x data	This study	CaltechData: https://doi.org/10.22002/D1.1915
processed SMARTseq data	This study	CaltechData: https://doi.org/10.22002/D1.1900
raw rheology video data	This study	CaltechData: https://doi.org/10.22002/D1.1905
transcriptome data	This study	CaltechData: https://doi.org/10.22002/D1.1914
RNAi experiments, survival assays, in vitro enzyme data	This study	CaltechData: https://doi.org/10.22002/D1.1917
alignment and tree fasta files	This study	CaltechData: https://doi.org/10.22002/D1.1916
genome and transcriptome assemblies of other rove beetles, including analysis script	This study	CaltechData: https://doi.org/10.22002/D1.1919
Drosophila melanogaster	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/215/GCF_000001215.4_Release_6_plus_ISO1_MT/GCF_000001215.4_Release_6_plus_ISO1_MT_protein.faa.gz

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Agrilus planipennis	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/699/045/GCF_000699045.2_Apla_2.0/GCF_000699045.2_Apla_2.0_protein.faa.gz
Anoplophora glabripennis	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/390/285/GCF_000390285.2_Agla_2.0/GCF_000390285.2_Agla_2.0_protein.faa.gz
Leptinotarsa decemlineata	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/500/325/GCF_000500325.1_Ldec_2.0/GCF_000500325.1_Ldec_2.0_protein.faa.gz
Dendroctonus ponderosae	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/355/655/GCF_000355655.1_DendPond_male_1.0/GCF_000355655.1_DendPond_male_1.0_protein.faa.gz
Aethina tumida	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/937/115/GCF_001937115.1_Atum_1.0/GCF_001937115.1_Atum_1.0_protein.faa.gz
Tribolium castaneum	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/335/GCF_000002335.3_Tcas5.2/GCF_000002335.3_Tcas5.2_protein.faa.gz
Onthophagus taurus	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/648/695/GCF_000648695.1_Otau_2.0/GCF_000648695.1_Otau_2.0_protein.faa.gz
Nicrophorus vespilloides	NCBI nr (protein database)	https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/412/225/GCF_001412225.1_Nicve_v1.0/GCF_001412225.1_Nicve_v1.0_protein.faa.gz
Coproporus sp.	Genome sequencing; this study	Sequence Read Archive: SRR4301367; CaltechData: https://doi.org/10.22002/D1.1919
Tachinus sp.	Genome sequencing; this study	Sequence Read Archive: SRR15992418; CaltechData: https://doi.org/10.22002/D1.1919
Aleochara bilineata	NCBI BioProject	NCBI BioProject: PRJNA378164; CaltechData: https://doi.org/10.22002/D1.1919
Drusilla canaliculata	Genome sequencing; this study	Sequence Read Archive: SRR5906249; CaltechData: https://doi.org/10.22002/D1.1919
Lissagria laeviuscula	RNA sequencing; this study	Sequence Read Archive: SRR15992419, SRR15992420; CaltechData: https://doi.org/10.22002/D1.1919

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Platyusa sonomae	RNA sequencing; this study	Sequence Read Archive: SRR15992410, SRR15992411, SRR15992412, SRR15992413; CaltechData: https://doi.org/10.22002/D1.1919
Dalotia coriaria reference genome assembly	http://131.215.78.39/genomebrowser/cgi-bin/hgGateway?genome=Dcor2&hubUrl=http://131.215.78.39/genomebrowser/evolution/parkergroup/Dalotia_coriaria/hub2.txt	N/A
Other		
ZB-5MS fused silica capillary column	Phenomenex	7HG-G010-11
pluriStrainer cell strainer	ImTec Diagnostics	43-50040-51

RESOURCE AVAILABILITY

Lead contact

Further information and requests for reagents and resources should be directed to and will be fulfilled by the Lead Contact, Joseph Parker (joep@caltech.edu).

Materials availability

Primers and generated dsRNAs are available via request from the Lead Contact.

Data and code availability

Raw sequence reads related to this manuscript have been deposited on NCBI under the BioProject 'RNaseq (10x and SMARTseq) of the tergal gland of *Dalotia coriaria*' (BioProject: PRJNA707010) and 'Genome and transcriptome sequencing of rove beetles (Staphylinidae)' (BioProject: PRJNA764224). All other data were uploaded to CaltechData: https://data.caltech.edu/records/1915 (processed scRNaseq 10x data), https://data.caltech.edu/records/1900 (processed SMARTseq data), https://data.caltech.edu/records/1905 (raw rheology video data), https://data.caltech.edu/records/1914 (transcriptome data of *Dalotia*), https://data.caltech.edu/records/1919 (genomes and transcriptomes of other rove beetles), https://data.caltech.edu/records/1917 (RNAi experiments, survival assays, *in vitro* enzyme data), and https://data.caltech.edu/records/1916 (alignment and tree fasta files). Detailed code for scRNaseq analyses with Seurat and cNMF; video analyses of rheology data; custom R scripts for SMARTseq analyses via sleuth, GOterm assignments and survival/toxicity data analyses can be found on CaltechData (https://data.caltech.edu/records/1918). All other statistical comparisons using ANOVAs, Kruskal-Wallis tests, U-tests and simple ordinations were done in Past 3.04 (Hammer et al., 2001).

METHOD DETAILS

Beetle husbandry

The Greenhouse Rove Beetle (*Dalotia coriaria*, Kraatz) strain used in this study was originally donated by Applied Bionomics (Canada) and inbred for nine generations. Beetles were kept as previously described (Parker et al., 2018) and fed with oat/poultry-rearing pellet-powder three times per week.

Microbial suppression

Late third instar larvae were fed with a mixture of sterilized oat/poultry-rearing pellet-powder with amoxicillin, streptomycin and tetracycline antibiotics (2.5% w/w for each) or pure sterile oat/poultry-rearing pellet-powder as a control. Beetles were single-housed in 5 cm Petri dishes with a thin layer of Plaster of Paris. Petri dishes were moisturized three times a week and fresh food was provided *ad libitum*. Ten days after adult eclosion, tergal gland contents were extracted in hexane and analyzed via GC-MS (see below). To assess the effect of antibiotic treatment on absolute numbers of bacteria associated with the beetles, bacterial 16S rRNA copy numbers determined by quantitative PCR (qPCR). Since many insects have bacteria on their cuticles, the beetles were surface washed in 5% (*v/v*) sodium dodecyl sulfate solution before bacterial quantification. For the control and antibiotic treatments, DNA was extracted from eight replicates of individual *Dalotia* using the Zymo Research Quick-DNA Miniprep Plus Kit according to manufacturer's instructions. For qPCR, we used universal eubacterial 16S rRNA gene primers (Univ16SRT-F: 5′-ACTCCTACGGGAGGCAGCAGT-3′; Univ16SRT-R: 5′-TATTACCGCGGCTGCTGGC-3′) (Clifford et al., 2012). For quality assessment of DNA extracts and standardization

of bacterial titers, qPCR with primers targeting host 28S rRNA was conducted simultaneously (D3A _F: 5′-GACCCGTCTTGAAAC ACGGA-3′; and D3B_R: 5′-TCGGAAGGAACCAGCTACTA-3′) (Litvaitis and Rohde, 1999). qPCR was performed on an Applied Biosystems 7300 Real Time PCR System in final reaction volumes of 25 μl, including the following components: 1 μl of DNA template, 2.5 μl of each primer (10 μM), 12.5 μl of autoclaved distilled H_2O, and 6.5 μl of Luna® Universal qPCR Master Mix (New England BioLabs). PCR conditions were: 95°C for 5 minutes, 40 cycles at 95°C for 10 s, 70°C for 15 s, and 72°C for 10 s. A melting curve analysis was performed by increasing the temperature from 60°C to 95°C within 20 min. Standard curves were established for host 28S and bacterial 16S using PCR product as templates. A Qubit fluorometer (Thermo Fisher) was used to measure DNA concentrations to calibrate standard curves. The ratio between absolute copy numbers of bacterial 16S and host 28S (= bacterial/host copy ratio) was used as a standardized measure of bacterial abundance per beetle sample and the difference between the antibiotic-treatment and control group beetles was assessed with a Mann-Whitney-U-test.

Biochemical tracer experiments and SIM mass spectrometry

Late third instar larvae were fed sterilized oat/poultry-rearing pellet-powder with amoxicillin, streptomycin and tetracycline antibiotics (2.5% w/w for each) plus 25% (*w/w*) stable isotope-labeled precursors: $^{13}C_6$-D-glucose, $^{13}C_6$-tyrosine (all > 99% enrichment, Sigma-Aldrich), D_3-methione, D_{23}-dodecanoid acid and $^{13}C_6$-4-hydroxybenzoic acid (98%–99% enrichment, Cambridge Isotope Laboratories, Inc.) as well as a control with sterile oat/poultry-rearing pellet-powder. Beetles were single-housed and glands of adults were extracted ten days after eclosion using hexane. Crude hexane extracts were analyzed with a GC-MS as described in detail below. Electron ionization mass spectra of characteristic fragment ions were monitored in single ion mode (SIM) and at 70 eV.

Artificial disarming and survival biotest against ants

We developed a protocol to artificially disarm *Dalotia*, creating beetles that lack the tergal gland secretion. Adult beetles were collected from laboratory stock populations and placed on a CO_2 fly pad; after the beetles were initially anesthetized the valve was closed and beetles could recover. Subsequently, we pulsed the beetles with low doses of CO_2 which initiated abdomen flexing and visible release of chemicals from the tergal gland. We repeated this cycle of anesthesia, recovery, and low pulses of CO_2 five times. To check success of the protocol, a subgroup of CO_2-treated beetles, as well as control animals form the stock population beetles were individually extracted in 70 μL hexane for 10 min and their glandular compounds profiled with GC-MS (see below). Control group beetle glands contained 6.3 ± 5.6 μg (mean ± SD) of secretion, while CO_2-treated animals contained 0.5 ± 1 ng (Figure S1A). CO_2-treatment did not affect survival of the beetles and after 72 h all 25 beetles in both the treated and control groups were still alive. For the survival biotest, ten *Dalotia* beetles were paired with five *Liometopum occidentale* Emery ants (collected: Chaney Canyon, Altadena, CA; 34°13'04.1"N 118°09'06.4"W). Beetles and ants were placed in a 100x100x50mm plastic box with 10mm Plaster of Paris and two rolled pieces of Kimwipe to create an artificial foraging space (Vucic-Pestic et al., 2010). In total, we prepared 14 boxes each of control and disarmed beetles (= 140 beetles and 70 ants per treatment). The experiment ran for 48 h and the percentage of surviving beetles calculated for each box. Difference in survival between control and disarmed groups was analyzed using a Mann-Whitney-U-test.

Double-stranded RNA preparation and RNAi knockdown in Dalotia

Double-stranded RNA was prepared from cDNA from pooled from all life-stages of the beetle. Regions of 450-650 bp for locus were amplified using primers with T7 linkers (5′- TAATACGACTCACTATAGGG-3′). Fragments were cloned into a pCR4-TOPO Vector (TOPO TA Cloning Kit, ThermoFisher) using primers listed in the Key Resource Table.

The same primers were subsequently used to amplify template DNA from the TOPO vector for dsRNA synthesis, using the MEGAscript T7 Transcription Kit (ThermoFisher). After synthesis, dsRNA was cleaned using acid phenol/chloroform (50:50) and adjusted to a concentration of ~6 mg/ml. For injection, dsRNA stock was then diluted 1:1 in DEPC-treated 1x PBS and blue food dye following a previously published protocol (Philip and Tomoyasu, 2012). Following the same protocol, dsRNA against green fluorescent protein (*GFP*) was prepared and injected as a control. Late third instar larvae were collected from laboratory stock populations and microinjection was performed according to Parker et al. (2018). After injection, the larvae were individually placed into 5 cm plastic Petri dishes with thick moistened filter paper. Larvae that died before pupation or did not pupate by the end of ten days were discarded. The filter paper was moistened three times per week. After adults had eclosed, beetles were additionally fed with frozen fruit flies and oat/poultry-rearing pellet-powder on the same schedule for ten days. Mature beetles were used for experiments chemical assays.

Extraction of gland compounds and gas chromatography/mass spectrometry

To screen chemical phenotypes of dsRNA knockdown and dsGFP control beetles, single specimens were submersed (Dettner, 1984, 1993b; Steidle and Dettner, 1993) in 70 μL hexane (company, GC/MS analytical grade) containing 150 ng/μl n-octadecane (Sigma Aldrich) as an internal standard; after 10 min the solvent was separated from the insect, transferred into a new vial and frozen at −80°C for further analysis. A GCMSQP2020 gas chromatography/mass-spectrometry system (Shimadzu, Kyōto, Japan) equipped with a ZB-5MS fused silica capillary column (30 m x 0.25 mm ID, df = 0.25 μm) from Phenomenex (Torrance, CA, USA) was used for chemical profiling. Crude hexane sample aliquots (2 μl) were injected by using an AOC-20i autosampler system (Shimadzu) into a split/splitless-injector which operated in splitless-mode at a temperature of 310°C. Helium was used as the carrier-gas with a constant flow rate of 2.13 ml/min. The chromatographic conditions were as follows: The column temperature at the start was 40°C with a 1-minute hold after which

the temperature was initially increased 30°C/min to 250°C and further increased 50°C/min to a final temperature of 320°C and held for 5 minutes. Electron impact ionization spectra were recorded at 70 eV ion source voltage, with a scan rate of 0.2 scans/sec from m/z 40 to 450. The ion source of the mass spectrometer and the transfer line were kept at 230°C and 320°C, respectively.

Compounds were identified based on their *m/z* fragmentation patters and by comparison to authentic standards (1,4-benzoquinone, 2-methyl-1,4-benzoquinone, undecane, isopropyl decanoate, ethyl decanoate, ethyl dodecanoate; all SigmaAldrich; 2-methoxy-3-methyl-1,4-benzoquninone was synthesized as outlined below). Additionally, compound identity was confirmed by comparison of retention indices and MS data with published literature (Dettner, 1984; Dettner et al., 1985; Steidle and Dettner, 1993). The authentic standards were used to construct four-point calibrations curves for external standardization and quantification of BQs, esters and alkanes. Semi-quantification of bulk cuticular hydrocarbons (CHCs) was based on the internal C18 standard. We quantified the ion abundance and the absolute amounts of compounds in ng based on the internal or external standard (see above), as well as the relative composition of individual CHCs compared to ion abundance of the other CHCs. For quantifying gland compounds following RNAi silencing, we compared the target compound amounts between dsRNA knockdown and dsGFP control beetles with a Kruskal-Wallis test or Mann-Whitney-U-tests. Compositional data of CHCs was compared by non-metric multidimensional scaling (NMDS) ordination of a Bray-Curtis similarity matrix and analyzed using PERMANOVA (Anderson, 2001; Brückner and Heethoff, 2017). GC-MS data related to RNAi can be found at CaltechData (https://data.caltech.edu/records/1917).

In situ hybridization chain reaction (HCR)

For *in situ* HCR, DNA probe sets, DNA HCR amplifier, HCR hairpins as well as hybridization, wash and amplification buffers were purchased from Molecular Instruments (Beckman Institute at Caltech; https://www.molecularinstruments.com/) for each target transcript (Key Resources Table). For dissections, adult beetles were immersed in ice-cold DEPC-treated PBS and their abdomens were removed. Ventral segments were removed with dissection scissors. Dorsal abdominal segments were fixed in 4% PFA in DEPC-treated PBST for 25 minutes at room temperature and subsequently washed with DEPC-treated PBST (3x5 min). Fixation was followed by a 5 min proteinase K (10 μg/mL) treatment at room temperature, and samples were then rinsed in DEPC-treated PBST (3x). Samples were postfixed in 4% PBST-PFA for 25 minutes at room temperature and washed again with DEPC-treated PBST (3x5 min). The amplification and detection stages followed published protocols (Choi et al., 2018). Probes were either initiated with B1-Alexa546 or B4-Alexa488 amplifiers. After amplification and before the final wash steps, DAPI or Hoechst 33342 (1:2000) to mark nuclei, Alexa 488- or Alexa-647-Wheat Germ Agglutinin Conjugate (WGA; 1:200) to label cell membranes and in selected samples BODIPY 493/503 (1:200) to stain for neutral lipids were added. Tissue samples were imaged as whole mounts of dorsal abdomens in ProLong Gold Antifade Mountant (ThermoFisher), using a Zeiss LSM 880 with Airyscan fast.

Fly toxicity biotest

To test the toxicity of *Dalotia* tergal gland compounds, we followed an approach by Dettner (1984). First, artificial test solutions were prepared using the main gland constituents mimicking the natural ratios. An all-compound mixture (~100 μl) was prepared by mixing 8 mg 1,4- benzoquinone, 28 mg 2-methyl-1,4-benzoquinone, 47 mg undecane, 7 mg isopropyl decanoate, 13 mg ethyl decanoate and 7 mg ethyl dodecanoate (all Sigma Aldrich). Subsequently, mixtures were prepared without certain compounds to create the following treatments: undecane+BQs, esters+BQs, all BQs, undecane+esters, undecane, esters; 1x PBS was used as control. For the survival tests a group of 25 *Drosophila melanogaster* third instar wandering larvae were immersed for 1 s in 1 mL of artificial secretion or dipped into solid BQ powder and subsequently moved into a fresh culture tube. This protocol was repeated five times for each of the eight different treatments. Numbers of surviving fly larvae were counted after 1 h. For statistical analysis we used a generalized linear model (GLM) with binomial error distribution and logit as link-function in R. We fitted 'fly survival' as a binomial response variable of surviving fly larvae and total number of fly larvae (25 individuals) in each replicate. Finally, we used the R-implemented 'Simultaneous Tests for General Linear Hypotheses' with Tukey contrast for pairwise comparisons.

Antimicrobial assay

To assay the antimicrobial properties of the *Dalotia* tergal gland secretion, we performed bacterial growth assays in a standard 96-well plate design. Artificial test solutions that mimicked natural ratios of compounds, as well and individual and pairs of compound classes were prepared as above (Fly toxicity biotest) and diluted in LB broth to a concentration of ~6 μg/μl. 1x PBS was used as a control. Cultures of *Pseudomonas fluorescens* wild-type (strain WCS365) (Jorth et al., 2019) were grown on LB agar plates overnight at 30°C. Subsequently, ~5 μl of the bacterial colony was transferred into 500 μl and the OD_{500} was measured on a NanoDrop Spectrometer (Thermo Fisher). Based on the OD reading, the cell suspension was diluted to $OD_{500} = 0.1$. For the growth assay, 15 μl of the cell suspension was inoculated with 134 μl LB and 1 μl of test secretion, and overlayed with 50 μl sterile mineral oil. Growth curves were recorded over 24h every 10 min on a Biotek Synergy 4 running in continuous shaking mode at 30°C. Each treatment and the negative PBS control were replicated 13 times across two plates.

The OD_{500} values were plotted over time, and we modeled the bacterial growth with a Bayesian regression using a normal likelihood with mean given by a logistic function:

$$OD_{500}(t) = \frac{A}{1 + e^{\frac{4\mu_m}{A}(\lambda - t) + 2}}$$

Where λ describes the lag time (in h) before bacterial growth takes off, μ_m is the maximal growth rate, t is the time since inoculation and A is the asymptote of the curve after growth levels off. We sampled these parameters for each growth event using the Stan implementation of Hamiltonian Monte Carlo and took the median sample value for each parameter for a given growth curve. A detailed script can be found on CaltechData (https://data.caltech.edu/records/1918). We then compared the lag time λ and the OD_{500} at the endpoint data after 24h for each treatment using a Welch's ANOVA and one-way ANOVA, respectively, followed by a Tukey's HDS pairwise posthoc test.

Ant selection experiment

To test for an adaptive role of the BQ and solvent cells in chemical defense, we generated beetles with either BQ-free or solvent-free gland secretions by knocking down *dmd* or *MFASN*, respectively, by injecting dsRNA (~3 mg/ml) against these targets. As a control treatment, we injected dsRNA against *GFP* (~3 mg/ml). Due to some mortality post-injection, we injected ~400-500 larvae for each dsRNA target. The detailed RNAi protocol, primer sequences and beetle care were the same as those used for pathway characterization (see below). 10-day old adults were used to set up the following survival assays: like previously, ten beetles were paired with five ants (collected: Chaney Canyon, Altadena, CA; 34°13'04.1"N 118°09'06.4"W) in a 100x15 mm plastic Petri dish with a thin layer of Plaster of Paris and two rolled pieces of Kimwipe. In total, we prepared 12 dishes per treatment (= 120 beetles and 60 ants per treatment), running the assay for 48 h before counted the number of surviving beetles. Because the amount of secretion per beetle cannot be determined before the experiment, we collected and extracted every beetle, regardless if dead or alive, in 30 μl of a mixture of 1:1 (v/v) hexane:chloroform for 15 min. We then ran a total of 360 beetle samples on the GC-MS using the same chromatographic conditions as outlined below. To determine if RNAi was successful in each individual beetle, we used single-ion detection of characteristic ions of 2-methyl-1,4-benzoquione (*m/z* = 122) and undecane (*m/z* = 156) to determine whether the glands of extracted beetles were benzoquinone-free (*dmd*-RNAi), solvent-free (*MFASN* RNAi) or not influenced (*GFP*). Only beetles with the correct gland composition were included in the final dataset; beetles with completely empty glands at the end of the experiment were removed from the dataset because it is possible they possessed empty glands at the start of the experiment. For statistical analysis of survival differences among BQ-free, solvent-free and control beetles, we again used a GLM with binomial error distribution and logit as link-function in R. We fitted 'beetle survival' as a binomial response variable of counted surviving beetles with a non-empty gland and total number of beetles (dead+alive) with a non-empty gland to account for the different sample sizes in each replicate. Finally, we used the R-implemented 'Simultaneous Tests for General Linear Hypotheses' with Tukey contrast for pairwise comparisons. The R script for this experiment can be found at CaltechData (https://data.caltech.edu/records/1918). As a control to assess survival without ants, we injected 50 larvae for each construct (*MFASN*, *dmd* and *GFP*), raised them until maturity and recorded their survival for a period of twelve consecutive days.

Bulk RNaseq and transcriptome assembly

For bulk RNA sequencing (3 biological and 3 technical replicates), a total of 105 gland (Segment A7) and control (Segment A6) segments from male beetles were dissected and separately collected for total RNA extraction with Trizol. Library preparation followed the Illumina TruSeq mRNA stranded kit protocol, including chemical shearing to obtain an average final library size of ~300 bp. Illumina 100 bp paired-end sequencing was performed on a HiSeq2500 platform in which the twelve *Dalotia* gland libraries were multiplexed on the same lane. For genome-guided assembly of the transcriptome, a bam-file was created from the *Dalotia coriaria* reference genome assembly: (http://131.215.78.39/genomebrowser/cgi-bin/hgGateway?genome=Dcor2&hubUrl=http://131.215.78.39/genomebrowser/evolution/parkergroup/Dalotia_coriaria/hub2.txt) using STAR (Dobin et al., 2013), while the approximately 240 million RNaseq reads were *in silico* normalized and subsequently used together with the bam-file to assemble transcripts using Trinity v2.8.4 (Grabherr et al., 2011). Finally, transcript sequences were clustered and concatenated to remove duplications using CD-HIT (Fu et al., 2012) with a sequence identity threshold of 0.95, yielding an assembly with a total length of 61.1 Mbp, an N50 = 3097 bp and a BUSCO score (Simão et al., 2015) of C:97.6%[S:67.5%,D:30.4%],F:0.5%,M:1.6%. Annotations of transcripts were transferred from the respective gene models from the genome and additionally confirmed by blastx (Altschul et al., 1990) of all transcripts against the non-redundant NCBI protein database (Pruitt et al., 2007). Finally, the bulk-transcriptome was used to build a kallisto (Bray et al., 2016) index file for pseudoalignment of single-cell RNaseq reads and cell-type specific, ultra-low input RNaseq (SMART-Seq). The assembled transcriptome can be found at CaltechData (https://data.caltech.edu/records/1914).

Cell-type specific transcriptome sequencing (SMART-seq)

For microdissection of specific gland cell types, male *Dalotia* beetles were injected with Invitrogen Wheat Germ Agglutinin, Alexa Fluor 488 Conjugate, to visualize the glandular structure *in vivo*, using a microinjector. Under fluorescence microscopy, we dissected the BQ cells in groups of 3-7 cells, the solvent cell tissue (~1000 cells) and entire tergite 6 (control tissue) in ice-cold DEPC-treated PBS. We used the NEBNext® Single Cell/Low Input RNA Library Prep Kit for Illumina® together with NEBNext® Multiplex Oligos for Illumina® (New England Biolab) for library preparation, including cell dissociation, reverse transcription of poly(A) RNA, amplification full-length cDNA, fragmentation, ligation and final library amplification according to the manufacturer's protocol. We performed cDNA amplification for 18-20 PCR cycles and final library amplification for 8 PCR cycles. In total, we constructed 31 libraries (ten D1, ten D2 and eleven control). The quality and concentration of the resulting libraries were assessed using the Qubit High Sensitivity dsDNA kit (Thermo Scientific) and Agilent Bioanalyzer High Sensitivity DNA assay. Libraries were sequenced on an Illumina

HiSeq2500 platform (single-end with read lengths of 50 nt) with ~20-25 million reads per library. Illumina sequencing reads were pseudoaligned to the bulk transcriptome and quantified (100 bootstrap samples) with kallisto 0.46.0 (Bray et al., 2016).

Cell-type specific transcriptome analysis

For cell-type specific differential expression analysis, kallisto-quantified RNaseq data were processed with sleuth 0.30.0 (Pimentel et al., 2017) using Likelihood Ratio tests in R 3.6.1 (R Core Team, 2018). To assign either up or downregulation for each transcript we extracted the normalized transcripts per million (tpm) from gland and control samples of sleuth object and defined up and downregulation as:

$$\text{down}-\text{regulated}: \frac{\frac{1}{n_g}\sum_{i=1}^{n_g} x_i}{\frac{1}{n_c}\sum_{j=1}^{n_c} x_j} < 1; \textit{up}-\textit{regulated}: \frac{\frac{1}{n_g}\sum_{i=1}^{n_g} x_i}{\frac{1}{n_c}\sum_{j=1}^{n_c} x_j} > 1$$

Where x_i is the transcripts per million of a gene y across all gland libraries (n_g), and x_j is the transcripts per million of a gene y across all control libraries (n_c). The code can be found in the R script 'sleuth_R_file' in the Script repository on CaltechData (https://data.caltech.edu/records/1918). Based on the differential expression data, we extracted transcripts that were down or upregulated in the BQ or solvent gland cells compared to control tissue.

Preparation of single cell suspension for 10x scRNaseq

One hundred and fifty tergal gland segments (Segment 7) and control segments (Segment 6) of *Dalotia* males were dissected in EBSS (Earle's Balanced Salt Solution; ThermoFisher) and transferred to into 150 μl ice-cold Schneider's *Drosophila* medium and fetal bovine serum (SDM+FBS, 90:10 V/V; both ThermoFisher). After centrifugation at 250x g for 2 min, subsequent washing in 300 μl ice-cold EBSS and centrifugation for another 2 min at 250x g, the supernatant was replaced by 150 μl of pre-warmed (37°C) trypsin (Corning 0.25% Trypsin, 0.1% EDTA; Fisher Scientific). Cells were dissociated at 37C for 25 min and the reaction was reinforced by pipette mixing every 5 min. After 25 min, trypsin was inhibited with 100 μl soybean trypsin inhibitor (SigmaAldrich) and incubated for 5 min. Subsequently, 500 μl SMD+FBS was added and the cell suspension was passed through a 40 μm cell strainer (pluriStrainer, ImTec Diagnostics). Finally, cells were pelleted at 300x g for 8 min and resuspended in Dulbecco's Phosphate-Buffered Saline (DPBS) 0.04% BSA (ThermoFisher) prior to 10x scRNaseq runs. For cell counting the LIVE/DEAD Viability/Cytotoxicity Kit (Invitrogen, ThermoFisher) was used. The above protocol was repeated twice for segment 6 and three times for segment 7, yielding five 10x runs.

Single-cell RNA sequencing (scRNaseq) library preparation and sequencing

After manually determining the cell concentration using a hemocytometer, suspensions were further diluted to desired concentrations (~1000 cells/μL) if necessary. The appropriate volume of reverse transcription mix was added to target 2,500–10,000 cells for loading into the 10x chip. The Chromium Single Cell 3′ Library & Gel Bead Kit v3, Chromium Single Cell 3′Chip kit v3 (PN-120236), and Chromium i7 Multiplex Kit were used for all downstream RT, cDNA amplification (12 PCR cycles), and library preparation as instructed by the manufacturer (Chromium Single Cell 3′Reagents Kits v3 User Guide). The quality and concentration of the resulting libraries were assessed using the Qubit High Sensitivity dsDNA kit (Thermo Scientific) and Agilent Bioanalyzer High Sensitivity DNA assay. scRNaseq libraries were sequenced on an Illumina HiSeq4000 or NovaSeq6000 (paired-end with read lengths of 150 nt) and reads were aligned to the reference transcriptome using kallisto 0.46.0 (Bray et al., 2016). From a total of 1,669,423,289 reads, 86.3% mapped to the *Dalotia* transcriptome using kallisto | bustools (Melsted et al., 2019, 2021), giving an approximate sequencing depth of 100,000 reads per cell. The total number of transcripts detected across all cells was 26,259 (58.3%) out of 45,044 RNAs in the transcriptome. The mean number of genes per cell ranged between 1,227 for Segment 7 and 1,219 for Segment 6 (CaltechData: https://data.caltech.edu/records/1915 and https://data.caltech.edu/records/1918).

Analysis of scRNaseq data

Kallisto alignment files were quantified with BUStools (v.0.39.3) (Melsted et al., 2021) and loaded into R using BUSpaRse (Moses and Pachter, 2021). First, we used knee plot methods as implemented in DropletUtils (Lun et al., 2019) to filter empty droplets before downstream analysis in Seurat v3.1 (Stuart et al., 2019). To create Seurat objects for each library, we filtered cells with low (nFeature_RNA > 200) or overly large (nFeature_RNA > 15,000) transcript counts, merged all objects into one for quality assessment, and then split by library for normalization and scaling. From three 10x runs for Segment 7 and two 10x runs for Segment 6, we obtained 10,364 high-quality cells after stringent filtering (4,663 and 5,701 cells from Segments 6 and 7, respectively). Individual transcript counts in each cell were normalized by the total number of counts for that cell, multiplied by a scale factor (10,000), and natural log-transformed (NormalizeData function). Expression of each transcript was scaled to achieve a mean of zero and variance of 1 across cells (ScaleData function). Transcripts with high overall expression and variance across cells were identified (FindVariableFeatures function; highest standardized variance was selected by selection.method = 'vst'). Subsequently, all five objects (libraries) were integrated, and shared sources of variation were identified by Seurat CCA alignment (Butler et al., 2018; Stuart et al., 2019). We selected the 3000 most variable transcripts across datasets (SelectIntegrationFeatures function) and used this set of transcripts to identify correspondences among the datasets, termed anchors (FindIntegrationAnchors function). Anchors were used to integrate the five datasets (IntegrateData function). The integrated Seurat object was again scaled and used as input for dimensionality reduction via Principal Component Analysis (PCA). Subsequently, we constructed a Shared Nearest Neighbor (SNN) graph for the dataset

(FindNeighbors function) using the first 20 principal components of the PCA as well as an SNN pruning cutoff of 0.067 before performing cell clustering with the Louvain algorithm (FindClusters function; resolution of 0.25-1.0). We eventually used a resolution of 0.55, as this resulted in the separation of cell clusters that we had previously identified as different cell types by HCR (see CaltechData: https://doi.org/10.22002/D1.1918). Finally, we ran the Uniform Manifold Approximation and Projection (UMAP) dimensional reduction technique on the 20 principal components using default options, creating the abdominal cell type atlas (Becht et al., 2018; McInnes et al., 2018).

Cell type annotation and statistical validation of cell classes

To annotate cell types, differentially expressed transcripts in cell clusters were identified in Seurat v3.1 (FindAllMarkers function) using both the Wilcoxon Rank Sum test as well as ROC analysis. Marker transcripts were computed using default settings for each cluster individually (FindMarkers function). A dendrogram was built in Past 4.02 (Hammer et al., 2001) using neighbor joining clustering (NJ) (see Musser et al. [2021] for a similar approach) on a Bray-Curtis similarity matrix (n = 10,000 bootstraps) based on the normalized and scaled cluster average expression profiles of the 3000 most variable transcripts. We used the list of overall differentially expressed genes, the individual cluster markers, the results of the NJ clustering and the transcript annotation, to connect cell types to known cell type classes using published marker genes from fly expression data using http://flybase.org/, https://www.uniprot.org/ and https://bgee.org (https://data.caltech.edu/records/1915).

To test the robustness of NJ clustering, we employed random forest analysis to assign cells to their pre-defined cell type classes. Random forest is a machine learning classifier that assigns samples (here cells) to predefined groups (here cell type classes) in multiple iterations and estimates the correct classification (Breiman, 2001). For the analysis, ntree = 1,000 bootstrap replicates were drawn with mtry = 55 variables randomly selected at each node. A confusion matrix showing the number of correctly assigned cell communities to either cell type as well as proportional class errors and total out-of-basket (OOB) estimate of error rates were computed.

Consensus non-Negative Matrix Factorization

We used the filtered cells-by-transcripts count matrix (10,634 cells by 26,259 transcripts) to infer gene expression programs (GEPs) of single cells using consensus Non-negative Matrix Factorization (cNMF) as implemented in Python (Kotliar et al., 2019). The method takes the count matrix (N cells x T transcripts) as input and produces both a gene expression program (GEP) x T matrix and a N x GEP matrix that specify the usage of each program for each cell in the data. The cNMF implementation runs multiple NMFs for each number of components (K) over a range of selected K, and combines the results of each replicate for different K to obtain more robust consensus estimates for the stability (silhouette score) and Frobenius error of each NMF model (Kotliar et al., 2019). We selected the most over-dispersed transcripts (T = 2000) using v-scoring (Klein et al., 2015) as the input for cNMF, which is essential for performing cNMF on biologically meaningful transcriptional variation (Kotliar et al., 2019). Expression of each over-dispersed transcript T was then scaled to unit variance and subjected to cNMF over a broad range of K (K = 16-36). The cNMF implementation was run with default parameters with the maximum number of iterations set to 250. Each iteration was run on the same normalized dataset over the range of Ks but with different randomly selected seeds. Subsequently, the factorizations for all iterations of each K were combined and the stability and error were computed. To choose the number of components K, we followed several approaches suggested by Kotliar et al. (2019); i) we plotted the stability and error as a function of K (Alexandrov et al., 2013), ii) performed a scree plot analysis of variance explained per principal component for the same dataset to confirm the appropriate choice of K (Eckart–Young–Mirsky theorem; Eckart and Young, 1936) and iii) re-ran the same cNMF steps with a variable number of over-dispersed T to see whether stability and error converge on a similar K (CaltechData: https://data.caltech.edu/records/1915 and https://data.caltech.edu/records/1918; Data S3).

Ultimately, we selected K = 20 and proceeded with the downstream cNMF workflow using default options, with density threshold r set to 0.05 in order to filter out replicates with low matching to a respective GEP (Data S3). We used the K = 20 cNMF dataset to calculate the normalized GEP usage score for each cell, and then loaded the cell type or cell class assignment from our previous Seurat clustering/NJ+random forest analysis into Python to calculate GEP usage (in %) for each cell type/class. The resulting GEP usage matrix for each cell type was visualized as a heatmap. We also exported the normalized cells-to-GEPs (N x GEP) and GEPs-to-transcript (GEP x T) matrices (CaltechData: https://data.caltech.edu/records/1915) and loaded the N x GEP matrix into R as a matrix, integrating the GEP usage data as a new assay into the pre-existing Seurat object (CreateAssayObject function). This enabled us to plot % GEP usage across single cells, permitting visualization of GEP usage across the UMAP cell atlas.

Finally, we extracted the expression data for the GEP9 to examine its composition (https://data.caltech.edu/records/1915). We assigned GO terms to each GEP9 transcript and performed over-representation analysis (Yu et al., 2012) of GO terms to visualize the number of genes for each GO-term after GO-slim. GO terms for the respective gene models were assigned based on the gene id with highest homology from the SwissProt database (download February 2019) or NCBI nr database (downloaded February 2019). A custom database of GO terms was created with makeOrgPackage function in the R package AnnotationForge v1.26.0 (Carlson and Pages, 2019). Over-representation analysis of GO terms was tested using the enrichGO function in the R package clusterProfiler v3.12.0 (Yu et al., 2012) with a hypergeometric distribution and a Fisher's Exact test. P values were adjusted for multiple comparisons using false discovery rate correction (Benjamini and Hochberg, 1995). Enriched GO-terms were subsequently visualized using the REViGO tool available at http://revigo.irb.hr and the MonaGO tool available at https://monago.erc.monash.edu/.

Genome assembly and gene prediction

Genomic DNA was isolated using a non-destructive extraction method described by Maruyama and Parker (2017) for *Coproporus* and *Drusilla canaliculata* or, for *Tachinus*, using the QIAGEN Blood & Cell Culture DNA Mini Kit following the manufacturer's protocol. DNA quantity was assessed using the Qubit High Sensitivity dsDNA kit (Thermo Scientific) and DNA integrity was assessed with Bioanalyzer. Illumina paired-end libraries (2 × 150bp) were prepared using the Illumina TruSeq DNA kit or NEBNext Ultra FS DNA library kit and sequenced on either the Illumina HiSeq X platform by Iridian Genomes/J. Parker and collaborators, or the HiSeq 2500 platform by the Millard and Muriel Jacobs Genetics and Genomics Laboratory at California Institute of Science and Technology. Reads for *Coproporus* and *Drusilla* were previously deposited on NCBI's Sequence read archive (*Coproporus* SRA id: SRR4301367; Drusilla SRA id: SRR5906249). Additionally, we downloaded reads of *Aleochara bilineata* (NICBI BioProject: PRJNA378164) (Kraaijeveld et al., 2019), re-assembled the genome and performed gene prediction.

Read quality for each taxon was assessed using FastQC v0.11.8 (Andrews, 2010). Illumina adapters, low-quality bases (phred scores below 15) and reads shorter than 50 bp were removed using cutadapt v1.18 (Martin, 2011). A preliminary assembly constructed from the filtered, adaptor-trimmed reads using MEGAHIT v1.1.3 (Li et al., 2016) with multiple *k-mers* (–k-list = 21, 29, 39, 59, 79, 99, 119) was filtered for bacterial contamination using Blobtools v1.0 (Laetsch and Blaxter, 2017). Contigs with sequence similarity to bacterial sequences to the NCBI nt database (downloaded February 2019, Evalue 1e-25) were removed. The filtered contigs were then assembled into scaffolds with three iterations of the Redundans v0.14a (Pryszcz and Gabaldón, 2016) reference-based pipeline using the *Dalotia* v1 genome assembly as the reference (–iters 3,–limit 0.5,–nogapclosing). Scaffolds smaller than 1 kb were removed and gaps filled using GapCloser v1.12 (Luo et al., 2012). Genome completeness was measured against the conserved Arthropoda v10 gene set (n = 1013 single-copy genes) using BUSCO v4.1.1 (Simão et al., 2015).

Species-specific repetitive elements were identified with RepeatModeler v 1.0.11 (Smit and Hubley, 2008) and MITE tracker (Crescente et al., 2018) as described by (Brückner et al., 2021). Repeats were further filtered for genuine gene products through sequence similarity searches against a local protein database of beetle sequences (RefSeq: GCF_001937115.1, GCF_000699045.2, GCF_000390285.2, GCF_000355655.1, GCF_000500325.1, GCF_001412225.1, GCF_000648695.1, GCA_001443705.1, GCF_000002335.3). Unclassified repeats with sequence similarity to known TEs were retained (E-value 1e-5) whereas those with no blast homology were removed (Petersen et al., 2019). Predicted repeats that were 80% identical were clustered with vsearch v 2.7.1 (–iddef 1–id 0.8–strand both) (Rognes et al., 2016). The final repeat library was used to soft mask the assemblies using RepeatMasker v 4.07 (Smit, 2004). For *Coproporus*, *Tachinus* and *A. bilineata*, a combination of *ab inito* (GeneMark-ES v.4.33 (Ter-Hovhannisyan et al., 2008) and Augustus v3.2.3 (Stanke and Waack, 2003) and reference proteome tools (GeMoMA v1.6.1 (Keilwagen et al., 2019)) were used to predict coding sequences. For *Drusilla*, in addition to the tools above, male and female whole body transcriptome evidence was used in two ways, through read mapping using BRAKER v2.1.2 (Brůna et al., 2021) and *de novo* Trinity assembled transcripts mapping using PASA v2.3.3 (Haas et al., 2008). Gene predictions were combined using EVidenceModeler v.1.1.1 (Haas et al., 2008) with the following weights for *Corproporus* and *Tachinus*: ab inito evidence = 1 and protein evidence = 5; and for *Drusilla*: ab inito evidence = 1, protein evidence = 1, PASA = 10, and BRAKER = 4.

RNA extraction and transcriptome assembly

Samples for RNaseq of *Lissagria laeviuscula* females and males (whole body) as well as *Platyusa sonomae* female and males (head and rest of the body separate) were placed in RNAlater and subsequently extracted following the Trizol protocol. Extracted RNA was shipped to Omega Bioservices (Norcross, GA, USA) on dry ice for library preparation and sequencing. The RNA for each sample was poly-A selected, cDNA was synthesized, and the final libraries were prepared following the Illumina TruSeq mRNA Stranded Kit protocol. Libraries were sequenced with ~90 million 150 bp paired-end reads on an Illumina HighSeq4000 platform. For each species, RNaseq reads were combined, *in silico* normalized and used for *de novo* transcriptome assembly with Trinity v2.8.4 (Grabherr et al., 2011). Sequences for each transcriptome were clustered and concatenated to remove duplications using CD-HIT (Fu et al., 2012) with a sequence identity threshold of 0.95. We then ran TransDecoder (Haas et al., 2013) to predict the longest ORFs and obtained preliminary annotations of transcripts by blastx search (Altschul et al., 1990) of all transcripts against the non-redundant NCBI protein database (Pruitt et al., 2007).

Phylogenetic analysis

To infer enzyme relationships, we first ran TransDecoder (Haas et al., 2013) using default settings to predict the longest ORFs in the bulk *Dalotia* transcriptome and translate them into protein sequences. We then ran cd-hit (Fu et al., 2012) with default options and a sequence identity threshold of 0.9 on this protein set to cluster them and remove short-length sequence artifacts. The protein sequences of eight previously published beetle genomes and *Drosophila melanogaster* were downloaded (see Key Resources Table) and combined with the predicted protein sequences from the newly generated genomic data for *Coproporus* and *Tachinus* (outgroup Tachyporinae) and *Drusilla canaliculata* (Aleocharinae: Lomechusini), as well as transcriptome data for *Lissagria laeviuscula* (Aleocharinae: Falagriini) and *Platyusa sonomae* (Aleocharinae: Lomechusini) (protein datasets listed below and deposited at CaltechData: https://data.caltech.edu/records/1919). We again ran cd-hit with default options and an identity threshold of 0.98 to remove isoforms. The protein sets were clustered into orthogroups with Orthofinder v2 (Emms and Kelly, 2015, 2019), using initial Diamond sequence alignment (Buchfink et al., 2015) and MCL clustering (Enright et al., 2002) to assign orthogroups. This was followed by multiple sequence alignment using MAFFT (Katoh and Standley, 2013) to align orthogroups and FastTree 2 (Price et al., 2010) to infer

Maximum Likelihood protein trees for each orthogroup. When then used the orthogroup alignments containing our target sequences and concatenated them into one master fasta-file for each enzyme family and realigned the protein sequences using MAFFT. All alignment files and tree files have been deposited at CaltechData (https://data.caltech.edu/records/1916). Master alignments were each analyzed using IQ-TREE v1.6.12 (Nguyen et al., 2015) with 1,000 bootstrap replicates. Where appropriate, enzyme trees were rooted using *D. melanogaster*. Trees were visualized with FigTree v1.4.4 (Rambaut, 2012). For most trees, we used *Drosophila* sequences as an outgroup. The laccase tree in Figure 6G is unrooted because the outgroup of insect laccases is unknown.

To characterize the cytochrome P450s of *Dalotia* belonging to the 4G subfamily, we downloaded a curated dataset of > 350 sequences by Feyereisen (2020) and added the two *Dalotia* paralogs, performing a phylogenetic analysis with the same parameters as described in Feyereisen (2020). The tree was rooted with a *Panonychus citri* CYP450 sequence.

Protein Expression and *in vitro* enzymatic assays for laccase

Synthetic cDNAs encoding the three laccases (Dmd, Lac1 and Lac2) were codon optimized for *E. coli* (CaltechData: https://data.caltech.edu/records/1917). Sequences were cloned into pTEV16 (VanDrisse and Escalante-Semerena, 2016) with a cleavable N-terminal 6xHis tag and TEV protease site. All proteins were expressed in BL21 competent *E. coli* cells (provided by C. VanDrisse, Caltech). Cells were grown in TB and protein expression was induced by 0.25 mM IPTG at OD600 $\approx$0.7 and harvested after 12 h at 20°C. Cells were lysed by freeze-thawing six times in lysis buffer (50 mM MES, 150 mM NaCl, 25 mM imidazole, DNase, and Thermo Scientific Halt Protease Inhibitor Cocktail). Proteins were passed over a column with Ni-NTA resin at room temperature, washed with 50 mM MES, 300 mM NaCl, 25 mM imidazole (pH = 7.5) and then eluted with 50 mM MES, 150 mM NaCl, 300 mM imidazole (pH = 7.5). Proteins were dialyzed with Thermo Scientific SnakeSkin for 4 h at room temperature into a final buffer of 10 mM MES and 100 mM $CuSO_4$ and then concentrated to 5 mg/ml for experiments and storage at −80C.

Enzymatic activity of each protein was tested against a standard substrate, ABTS (Data S6A). The reaction mixture was prepared as 5 mM MES, 0.3 M $CuSO_4$, and 2 mM of ABTS, with 2 μM of laccase. The UV spectrum was traced for three minutes (Data S6B). Subsequently, reactions were quenched with 5 M urea. To verify the RNAi results of the gland laccase Dmd, we set up the following reaction: 2mM of each HQ (1,4-hydroquinone, 2-methyl-1,4-hydroquinone and 2-methoxy-3-methy-1,4-hydroquinone) was separately added to vials containing the same reaction buffer previously used. A pilot experiment was set up for each compound and Dmd protein was added (same concentration as above) (Data S6C). this experiment was run for 10 min to determine the quenching time for the enzyme reactions via UV recordings every 30 s. Subsequently, 20 samples for each compound were prepared using the same method. To half of the samples we added 0.5 mM Dmd, while the other half served as control. Reactions were UV recorded for 1 minute and directly quenched with 0.05 M EDTA before being flash frozen and stored at −80C until further analysis. For the comparative analysis of Dmd, Laccase 1 and Laccase 2 we used the same buffers and substrates as previously. A total of three replicates and the same number of controls for each substrate-enzyme combination (i.e., 36 samples in total) were assayed.

To prepare samples for chemical profiling with either GC-FID or GC/MS for the Dmd versus control and comparative laccase assay, respectively, we used the following liquid-liquid extraction protocol. Frozen reaction mixtures were thawed and transferred into a clean glass vial with a sealed Teflon lid. For extraction, the aqueous phase was washed twice with 100 μl freshly Na_2SO_4-dried hexane. Organic fractions were combined, dried over Na_2SO_4 and carefully concentrated under a gentle stream of N_2 to a volume of $\approx$50 μl. For chemical analyses sample aliquots (1 μl) were injected by an AOC-20i autosampler system from Shimadzu, into split/splitless-injector which operated in splitless-mode at a temperature of 310°C. Either a GCMSQP2020 gas chromatography/mass-spectrometry system or a GC-2010PLUS with a flame-ionization detector (all Shimadzu, Kyōto, Japan) both equipped with a ZB-5MS fused silica capillary column (30 m x 0.25 mm ID, df = 0.25 μm) from Phenomenex (Torrance, CA, USA) were used. Helium was used as the carrier-gas with a constant flow rate of 2.13 ml/min. The chromatographic conditions were as follows: column temperature at the start was 40°C with a 1-minute hold after which the temperature was initially increased 50°C/min to 250°C and further increased 60°C/min to a final temperature of 300°C and held for 2 minutes. For the MS, electron impact ionization spectra were recorded at 70 eV ion source voltage, with a scan rate of 0.2 scans/sec from m/z 40 to 450. The ion source of the mass spectrometer and the transfer line were kept at 230°C and 320°C, respectively. For the FID, detector temperature was kept at 350°C. For quantification of the endpoint data (in ng) we used synthetic 1,4-BQ, 2-methyl-1,4-BQ (Sigma-Aldrich) and freshly synthesized 2-methoxy-3-methyl-1,4-BQ (see below).

Rheological measurements

To measure the surface-coating ability and extensional viscosity of different artificial mixtures of synthetic secretions, we used dripping-onto-substrate (DoS) rheology. In contrast to conventional rheology which requires sample sizes on the order of milliliters, a DoS measurement can be accomplished with microliters. The instrument comprises a controlled dispensing system, a nozzle and a high-speed camera (Dinic et al., 2017; Marshall et al., 2017; Rosello et al., 2019; Sousa et al., 2017). Mixtures were prepared using 1,4-BQ, 2-methyl-1,4-BQ, undecane, isopropyl decanoate, ethyl decanoate and ethyl dodecanoate (all Sigma Aldrich). 2-Methoxy-3-methyl-1,4-BQ was synthesized as follows: 2 mmol 2-Methoxy-3-methyl-1,4-hydroquinone (Sigma Aldrich) and 2 mmol $NaIO_4$ were mixed in a 50 mL round-bottom flask and 10 mL CH_2Cl_2 and 5 mL of water were then added. The biphasic mixture was stirred at room temperature for 10 min and the layers were subsequently separated and then the aqueous layer was washed with 10 mL CH_2Cl_2. All organic fractions were combined, dried over Na_2SO_4 and concentrated *in vacuo*.

The all-compound mixture mimicking the beetles' secretion was prepared by mixing 8 mg 1,4-benzoquinone, 18 mg 2-methyl-1,4-benzoquinone, 10 mg 2-methoxy-3-methyl-1,4-benzoquinone, 47 mg undecane, 7 mg isopropyl decanoate, 13 mg ethyl decanoate and 7 mg ethyl dodecanoate. For some experimental mixtures, undecane was replaced by tridecane, pentadecane, heptadecane or heptacosane (all Sigma Aldrich). The ratios of compounds in all mixtures always represented the average ratios of the respective compounds in *Dalotia*'s gland secretion (Figure 1D). Mixtures were dispensed at a rate of 0.02 ml/min through the nozzle with an inner diameter of 0.413 mm and outer diameter of 0.718 mm onto an aluminum substrate a distance of 2.5 mm below the nozzle. The fluid transitions from the nozzle to substrate via an unstable liquid bridge. The behavior of this liquid bridge was captured by the camera at 20,000 to 40,000 frames per second and analyzed to produce rheological data on the fluid (CaltechData: https://data.caltech.edu/records/1905 and https://data.caltech.edu/records/1918). Each measurement was repeated at least 3 times and the videos were processed using Python 3. Briefly, the videos were converted to a series of binary images and the minimum diameter at each frame was recorded. The code and parameters for each video are available in the supplement (CaltechData: https://data.caltech.edu/records/1918). This yielded a dataset of diameter and time which was fit to an equation describing power-law fluids:

$$\frac{R}{R_0} = Y(t_p - t)^n$$

Here, R_0 is 0.359 mm, n is the flow behavior index, and t_p is the pinch-off time, the moment the liquid bridge separates from the substrate. This equation was solved to estimate the elongational-flow consistency index Y ($1/s^n$) and the flow behavior index n. Y describes what happens if a flow in imposed on fluid, while n describes how Y changes as the flow rate changes. Y decreases as the ratio of surface tension to viscosity decreases. When n = 1, mixtures behave as Newtonian fluids (no change with flow rate) and when $n \neq 1$ they behave as non-Newtonian fluid or power law fluids. For $n < 1$, fluid flows more easily (shear thinning) as the flow rate increases. The closer n is to 1, the weaker is the dependence on flow rate. For $n > 1$, the fluid more strongly resists flowing (shear thickening). Fluids with both low Y and low n possess higher extensional viscosity (EV). Furthermore, previous work has shown $Y = \phi(n)\sigma/K$ (Doshi et al., 2003). This is the product of the shape factor, $\phi(n)$, and surface tension divided by the power-law flow consistency index (Blair et al., 1939). The data fit this power-law function from the last frame with $R/R_0 = 1$ to $t = t_p$, with 0 to 5 datapoints (a maximum of $\sim$125 μs) excluded near t_p because during this period the fluid deviates from the power-law behavior. We solved the previous equation to get K/σ which is the ratio of the flow consistency index and surface tension. We interpret this ratio as a metric of how a fluid flows over a surface at a given surface tension. K/σ can therefore be used as a proxy to describe the ability of a mixture to coat a surface, like the cuticle of other insects. We refer to this metric as the surface coating ability (SCA).

DATA ANALYSIS AND VISUALIZATION SOFTWARE

Analysis and visualization of transcriptomic data were performed using Python 3.0 (Van Rossum and Drake, 2000) and R v.3.6.0 (R Core Team, 2018) assisted by JupyterLab (Kluyver et al., 2016). The following R packages were used: AnnotationForge (Carlson and Pages, 2019), AnnotationHub (Morgan et al., 2019), biomart (Durinck et al., 2009) boot (Canty and Ripley, 2017), BUSpaRse (Moses and Pachter, 2021) car (Fox et al., 2012), clusterProfiler (Yu et al., 2012), cowplot (Wilke, 2017), data.table (Dowle et al., 2019), dbplyr (Wickham and Ruiz, 2019), DropletUtils (Lun et al., 2019) effects (Fox, 2003), ggplot2 (Wickham, 2011), lattice (Sarkar, 2008), lme4 (Bates et al., 2007), MASS (Ripley et al., 2013) Matrix (Bates and Maechler, 2010), nlme (Pinheiro and Bates, 2007), randomForest (Liaw and Wiener, 2002), Seurat v3.1 (Stuart et al., 2019) sleuth (Pimentel et al., 2017), tidyverse (Wickham et al., 2019), vcd (Meyer et al., 2020).

Supplemental figures

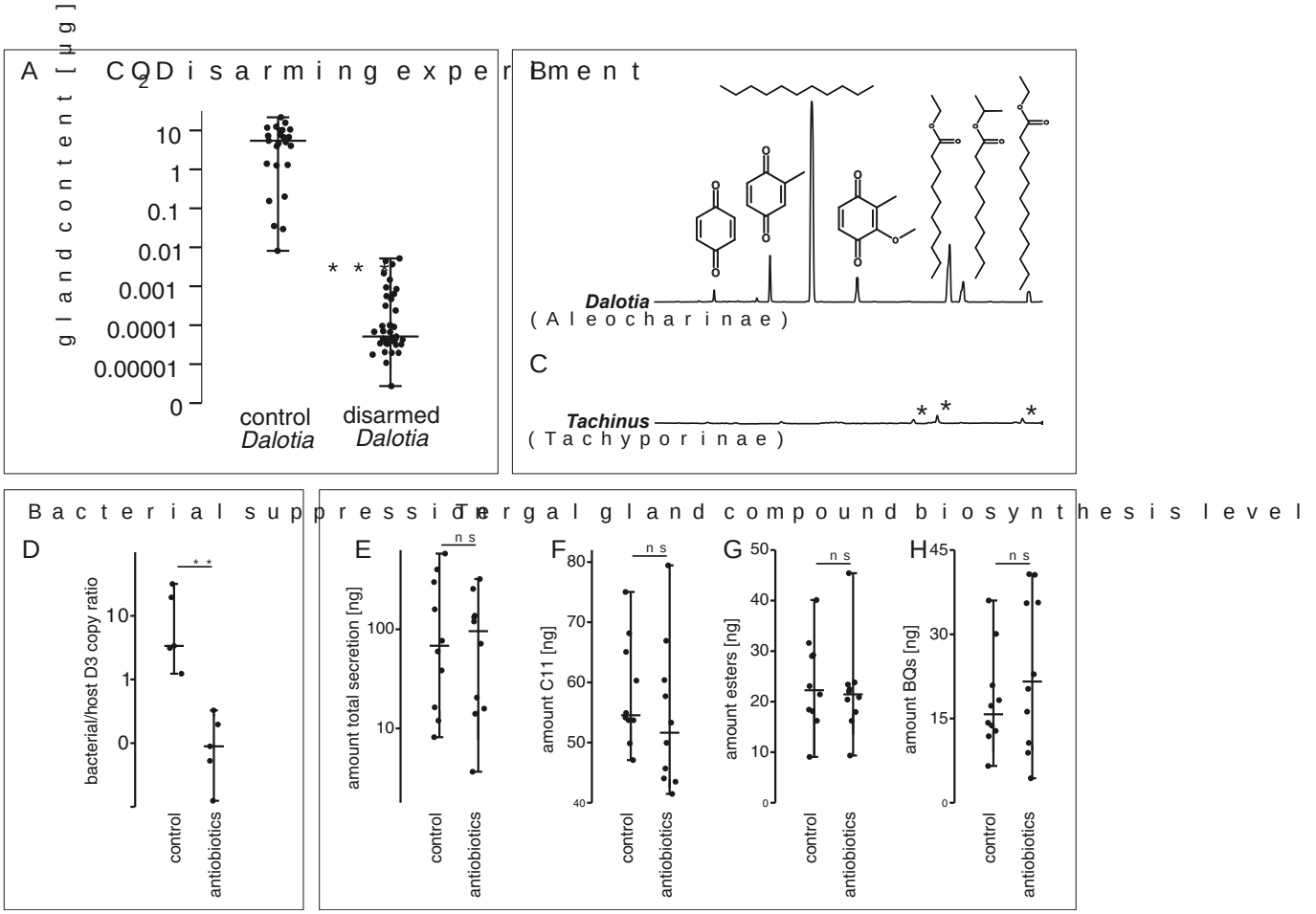
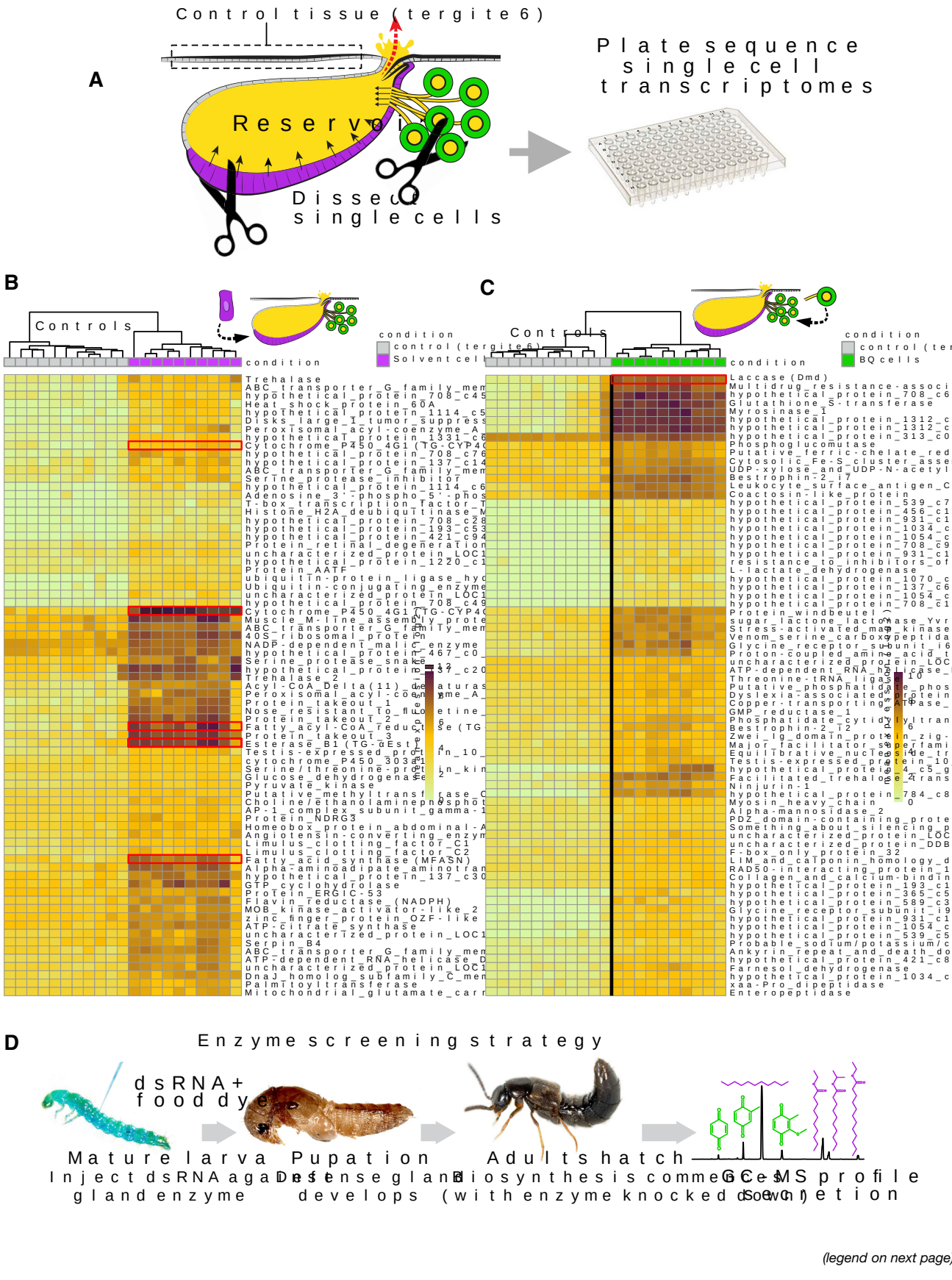


Figure S1. CO_2 disarming, comparison to a glandless outgroup rove beetle, and compound biosynthesis despite microbial suppression, related to Figures 1 and 2

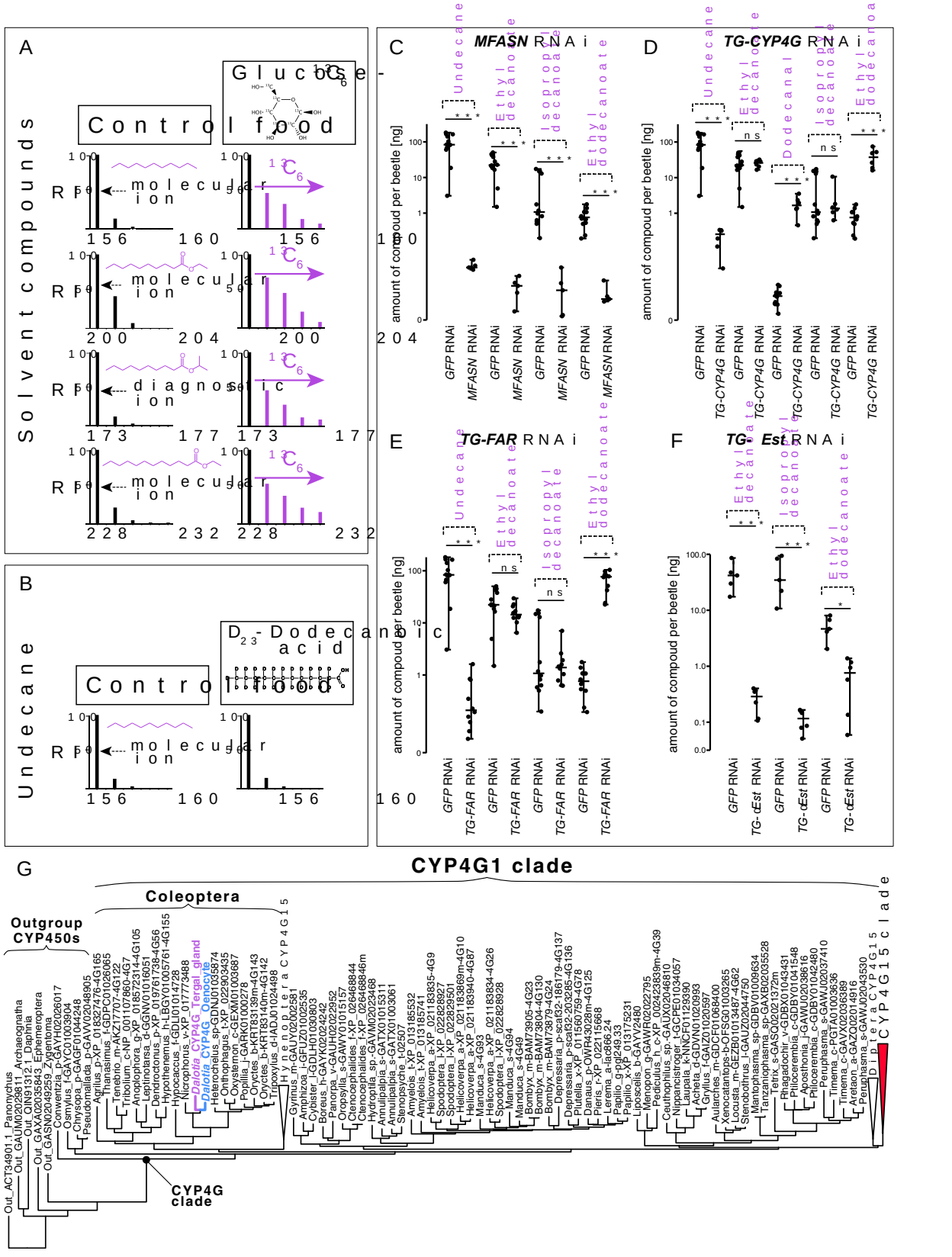
A: Repeated treatment with small doses of CO_2 leads to depletion of the tergal gland content in *Dalotia* effectively disarming the beetles. Significance was assessed using a Mann-Whitney-U test (*** = $p < 0.001$). B: Representative GC profile of the *Dalotia coriaria* gland content extracted in hexane. In order of retention: 1,4-benzoquinone (1,4-BQ), 2-methyl-1,4-BQ, undecane, 2-methoxy-3-methyl-1,4-BQ, ethyl decanoate, isopropyl decanoate, ethyl dodecanoate. C: Representative GC profile of *Tachinus* (Tachyporinae) extracted and measured using the same protocol. No gland compounds were detected. Asterisks denote in order of retention: C16:1-fatty acid (FA), C16-FA, C18-FA. D: Antibiotic treatment eliminates bacterial load in *Dalotia* as determined by qPCR quantification of the ratio of bacterial 16 s rRNA to host 28 s rRNA levels in antibiotic-treated and control beetles. E-H: Quantities of defensive compounds are unaffected by bacterial suppression: total secretion (E); undecane (F); the ester fraction (G); and the benzoquinones (H). Significance was assessed using Mann-Whitney-U tests (** = $p \leq 0.01$; ns = $p \geq 0.05$). In all figures, the horizontal middle bars represents the median, while the vertical bars show the minimum and maximum.



(legend on next page)

Figure S2. Cell-type specific transcriptomes and functional screening pipeline, related to Figure 1

A: To prepare cell-type specific transcriptomes of the two tergal gland cell types, cells were micro-dissected by hand and segment 6 tergite tissue was used as a control. Cells were lysed, and libraries prepared and sequenced to perform differential expression analysis. B: Differential expression analysis of solvent cells versus control cells. Heatmap showing the 75 most upregulated transcripts in the solvent cells (as assessed by q-value using sleuth; see https://data.caltech.edu/records/1918). Expression in all individual sequencing libraries is shown. Darker squares denote a stronger mean expression value. A full list of transcripts can be found at CaltechData (https://data.caltech.edu/records/1900). C: Differential expression analysis of BQ cells versus control cells. Heatmap showing the 75 most upregulated transcripts in the benzoquinone cells (as assessed by q-value using sleuth; see https://data.caltech.edu/records/1918). Expression in all individual sequencing libraries is shown. Darker squares denote a stronger mean expression value. A full list of transcripts can be found at CaltechData (https://data.caltech.edu/records/1900). D: Schematic workflow of the enzyme screening strategy. By injecting mature larvae with double stranded RNA (dsRNA) targeting specific mRNAs, transcription is silenced in gland cells with temporal control, during pupal and early adult life when defensive compound biosynthesis occurs. The chemical composition of the tergal gland secretion can then be quantified in the mature adult via gas chromatography-mass spectrometry (GC-MS).



(legend on next page)

Figure S3. Undecane and ester biosynthesis, related to Figure 2

A: Representative mass spectra of molecular ion region of undecane, ethyl decanoate, isopropyl decanoate and ethyl dodecanoate from beetles fed with unlabeled *Dalotia* diet (control) or infused with D-$^{13}C_6$-glucose. Spectra recorded in single-ion mode; enriched ions were detected in all four spectra in beetles fed D-$^{13}C_6$-glucose. B: Representative mass spectra of molecular ion region of undecane from beetles fed with unlabeled *Dalotia* diet (control) or food infused with D_{23}-dodecanoic acid. Spectra recorded in single-ion mode; no enriched ions were detected. C-F: RNAi silencing of solvent pathway enzymes. C: Amounts of undecane, ethyl decanoate, isopropyl decanoate and ethyl dodecanoate extracted from beetles injected with dsRNA against *GFP* (control) and *MFASN*. Silencing expression of *MFASN* resulted in near- complete loss of solvents (undecane and esters). D: Amounts of undecane, ethyl decanoate, dodecanal, isopropyl decanoate and ethyl dodecanoate extracted from beetles injected with dsRNA against *GFP* (control) and *TG-CYP4G*. Silencing *TG-CYP4G* caused loss of undecane, but a large increase in both dodecanal and ethyl dodecanoate without influencing levels of ethyl decanoate and isopropyl decanoate. E: Amounts of undecane, ethyl decanoate, isopropyl decanoate and ethyl dodecanoate extracted from beetles injected dsRNA against *GFP* (control) and *TG-FAR*. Silencing TG-FAR resulted in loss of undecane but a large increase in ethyl dodecanoate without influencing levels of ethyl decanoate and isopropyl decanoate. F: Amounts of ethyl decanoate, isopropyl decanoate and ethyl dodecanoate extracted from beetles injected with dsRNA against *GFP* (control) and *TG-αEst*. Silencing TG-αEst resulted in near complete loss of all esters. In all panels, significance was assessed using Mann-Whitney-U tests (*** = $p \leq 0.001$; * = $p \leq 0.05$; ns = $p \geq 0.05$). In (C–F), the horizontal middle bars represents the median, while the vertical bars show the minimum and maximum. G: Maximum-likelihood phylogeny of the CYP4G gene family (best model: LG+R8). Some clades have been collapsed for presentation purposes. The topology confirms that both *Dalotia CYP4G* copies are the product of a recent duplication within the CYP4G1 clade, known for their role in cuticular hydrocarbon synthesis in insects. *Dalotia* lacks a member of the CYP4G15 clade (red collapsed clade).

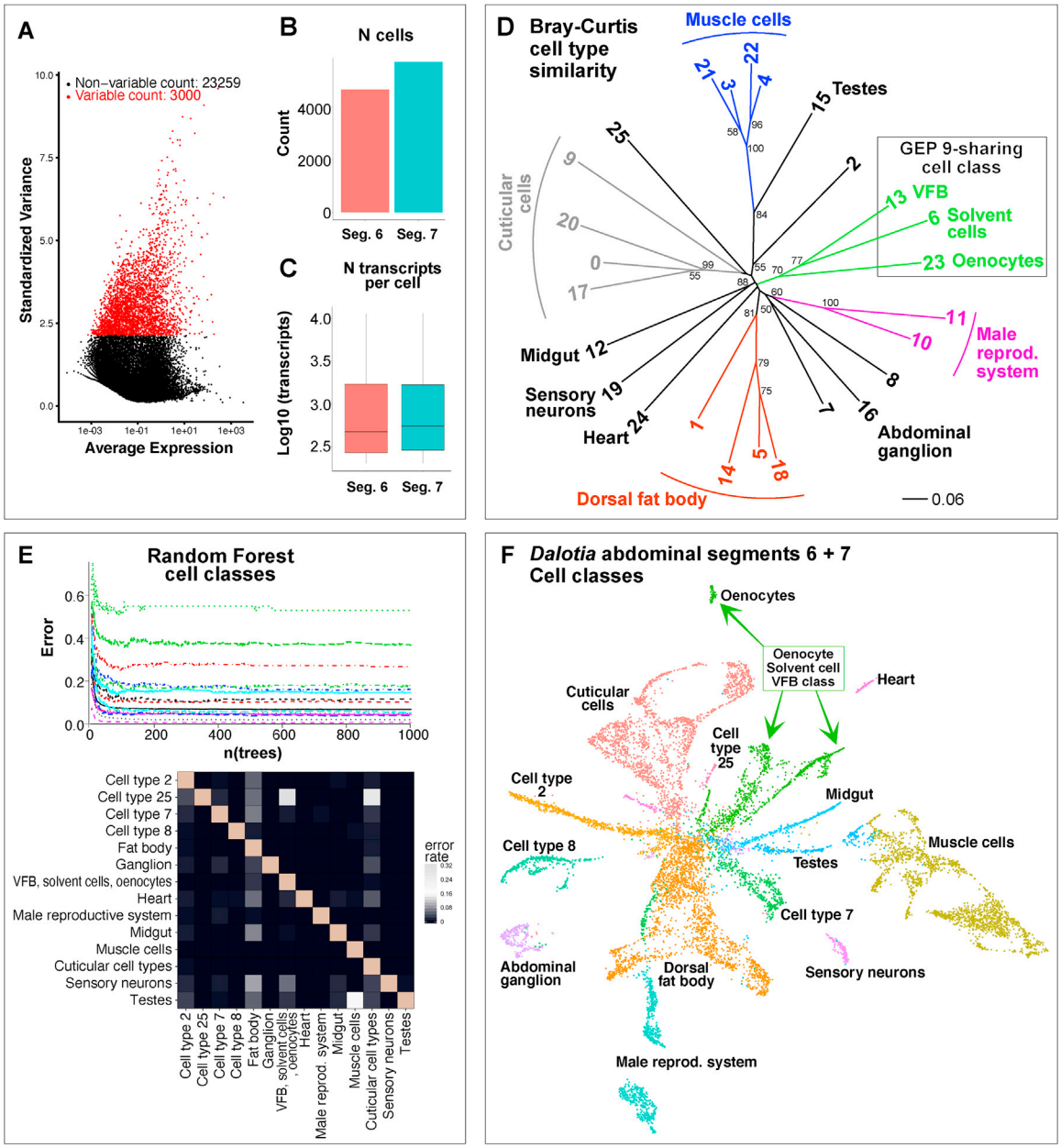
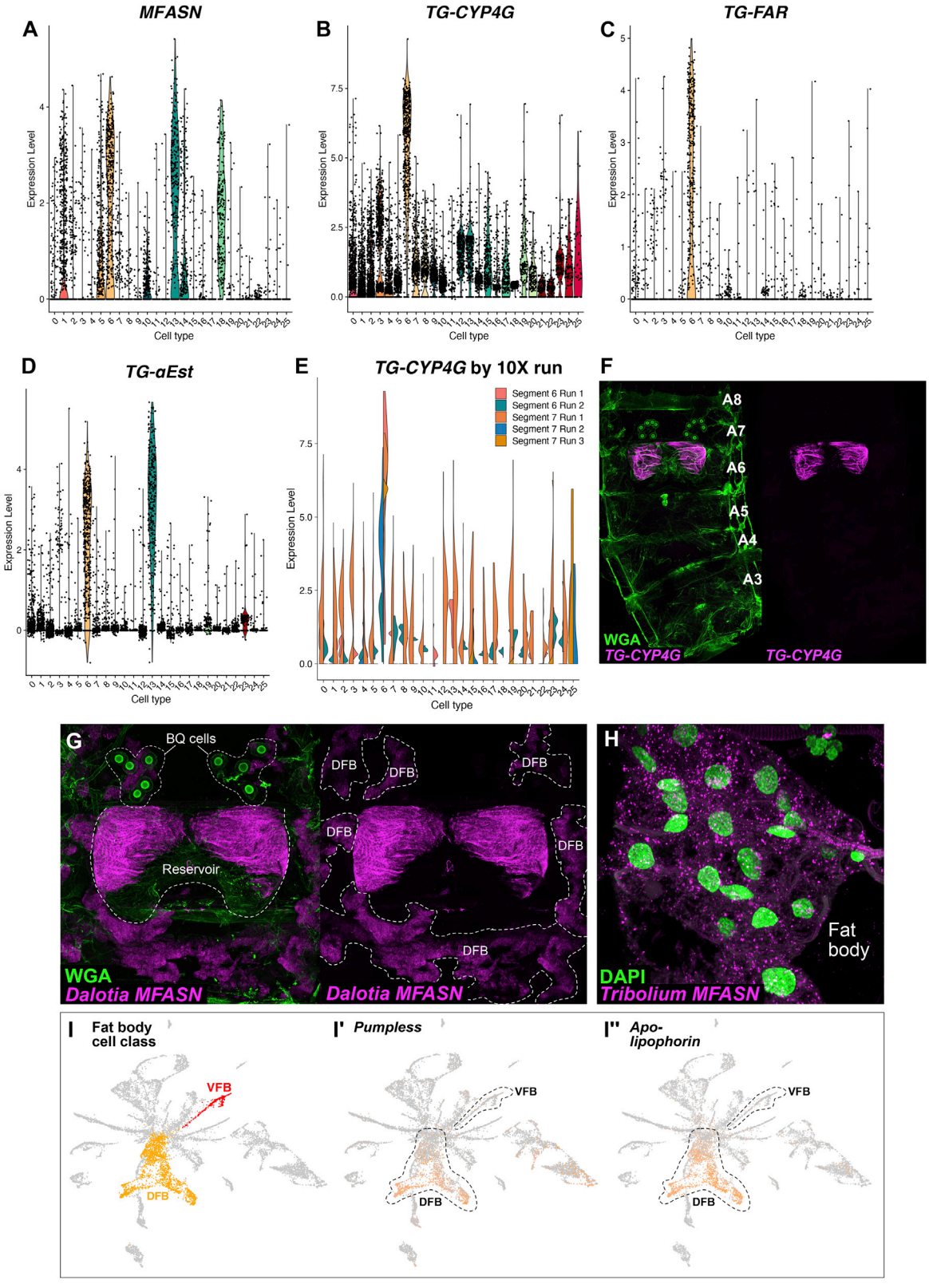


Figure S4. scRNA-seq data processing and cell clustering and classification, related to Figure 3

A: Variable feature plot showing the 3,000 transcripts that exhibit high cell-to-cell variation across the dataset (i.e., highly expressed in some cells and not others). These transcripts were used for all downstream analyses and clustering. B, C: The total number of cells (B) and number of transcripts per cell (C) from Segment 6 (control) and 7 (gland) of all concatenated 10x runs after filtering. In C, the middle bars of the boxplots represent the median, the boxes denote the interquartile range, while the whiskers show the minimum and maximum. D: Neighbor joining clustering (NJ) on a Bray-Curtis similarity matrix (n = 10,000 bootstraps) based on the normalized and scaled cluster average expression profiles of the 3000 most variable transcripts (A). Clades form the basis of further grouping the cell types into 14 higher order cell classes. E: Robustness of the 14 cell class assignments tested using a Random Forest classifier. The graph (top) shows the robustness (class error rate) of the cell class assignments running a Random Forest classifier over 1000 generations. The heatmap shows the pairwise-class error rates of the assigned groups after 1000 generations. Only cell classes with a relatively low cell number were observed to have a higher error rate. F: UMAP plot of *Dalotia* abdominal Segment 6 and 7 cell atlas showing 14 cell classes.



(legend on next page)

Figure S5. Solvent cell evolution, related to Figures 4 and 5

A-D: Violin plots showing expression of *MFASN* (A), *TG-CYP4G* (B), *TG-FAR* (C) and *TG-αEST* (D) in each cell across all cell types in the abdominal segment cell type atlas. E: Expression of *TG-CYP4G* by cell type in each of the five segment-specific 10x runs. *TG-CYP4G* shows a generally high expression across all cell types, but almost all expression in non-solvent cells comes from Segment 6 Run 1 and Segment 7 Run 1. This global expression, with very low levels in almost all cell types, is likely the result of the 'ambient mRNA'-effect, which is typical of droplet-based methods. Ambient mRNAs are most likely derived from high-abundance transcripts that are leaked from dying cells and bind to beads during the microfluidic step in the 10x workflow. Only in solvent cells are the UMI count levels were high, indicating that the low levels mRNA detected in non-solvent cells likely represent ambient mRNAs. F: To independently evaluate the cell type specificity of *TG-CYP4G*, we used HCR and imaged the whole beetle abdomen, confirming that its transcription is highly specific to the solvent cells. G: HCR in *Dalotia* reveals *MFASN* in the solvent cells (left panel) and dorsal fat body (DFB; right panel). C: HCR of the *MFASN* ortholog in *T. castaneum* reveals strong expression in fat body. H: UMAP plot of *Dalotia* abdominal segment 6 and 7 cell atlas highlighting the dorsal (DFB; orange) and ventral fat body (VFB; red) cells. I', I'': Expression of the *Pumpless* (I') and *Apolipophorin* (I'') fat body marker genes show strong expression in DFB but not VFB.

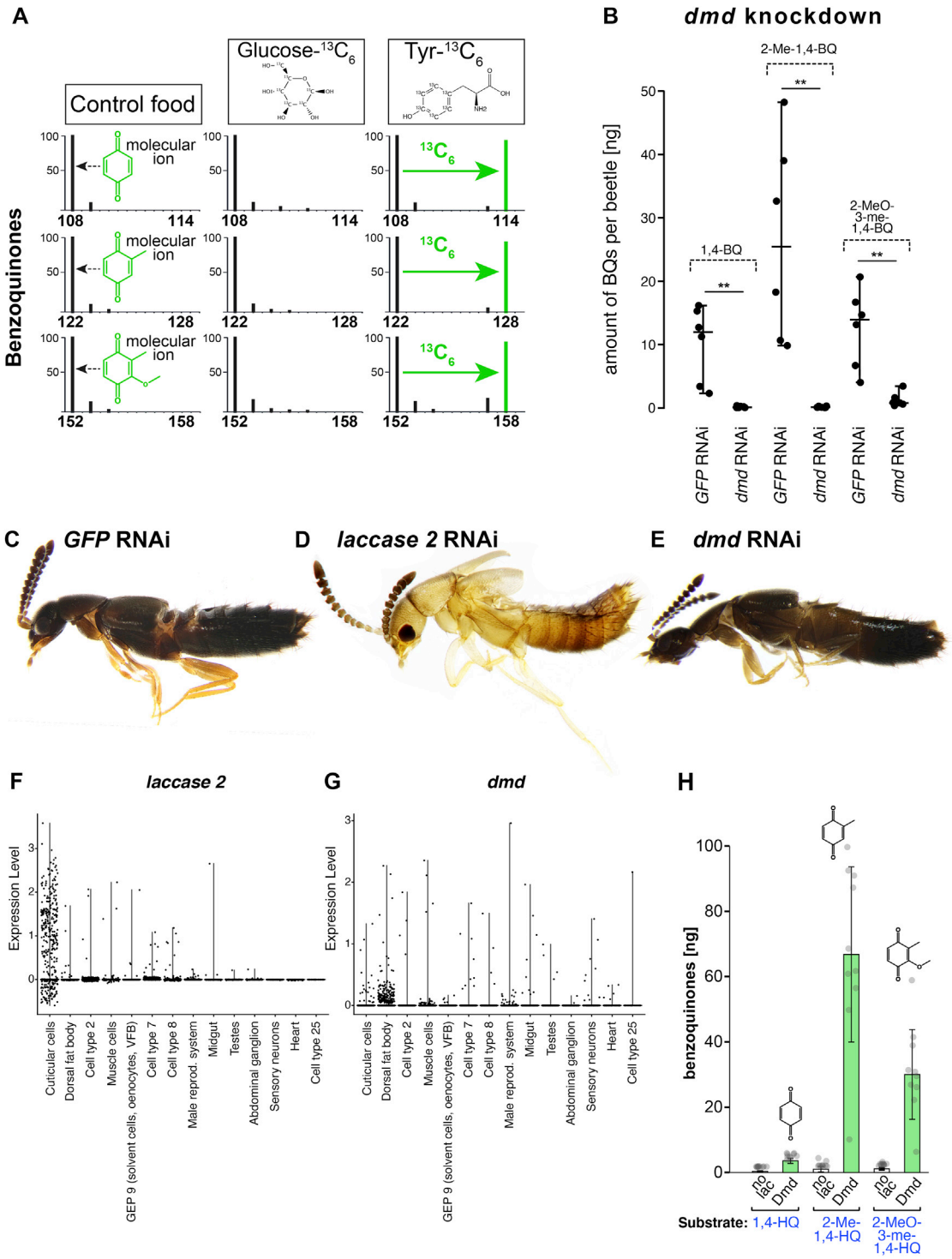


Figure S6. BQ cell evolution, related to Figure 6

A: BQs are synthesized from dietary Tyr but not from glucose. Representative mass spectra of molecular ion region of 1,4-BQ, 2-methl-1,4-BQ and 2-methoxy-3-methyl-1,4,-BQ from beetles fed with unlabeled *Dalotia* diet (control), food infused with $^{13}C_6$-glucose or $^{13}C_6$-Tyr. Spectra were recorded in single-ion mode. No enrichment was detected in any of the three BQs for the $^{13}C_6$-glucose fed beetles. In contrast, a strong $[M+6]^+$ enrichment was detected for all BQs in $^{13}C_6$-Tyr fed beetles. B: Functional characterization of Dmd. Levels of 1,4-BQ, 2-methyl-1,4-BQ and 2-methoxy-3-methyl-1,4,-BQ extracted from beetles injected with dsRNA against *GFP* (control) and *dmd*. Silencing *dmd* resulted in near-complete loss of BQs in *Dalotia*'s secretion. The horizontal middle bars represents the median,

(legend continued on next page)

while the vertical bars show the minimum and maximum. C-E: Adult beetle morphology after late larval injection with dsRNA against *GFP* (C); control), *lac2* (D) and *dmd* (E). (F): Expression of *laccase-2* across all cell classes in the 10x data reveals strong expression cuticle cells. (G): Expression of *dmd* across all cell classes in the 10x data shows no tissue specific expression of *dmd* as BQ-cells are not captured by 10x. In F and G, the horizontal middle bars represents the median, while the vertical bars show the minimum and maximum. H: Reaction of the three HQ substrates 1,4-HQ, 2-Me-1,4-HQ and 2-MeO-3-Me-1,4-HQ with purified Dmd *in vitro* as compared to a non-enzyme control. BQ levels were measured using GC-FID. Size of the bars represent the median, while error bars denote the SE.

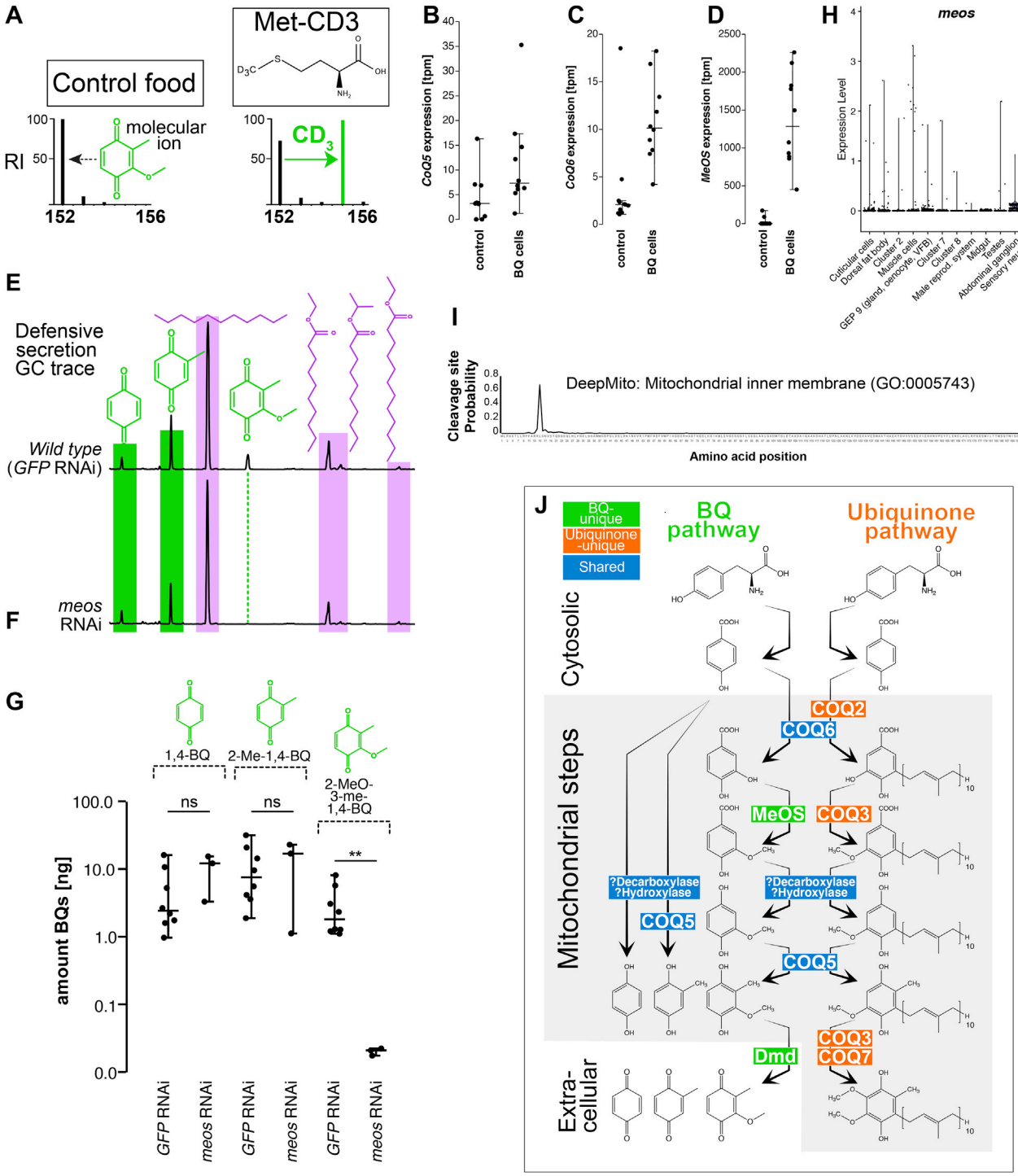


Figure S7. Evidence for a mitochondrial HQ pathway, related to Figure 6

A: Representative mass spectra of molecular ion region of 2-methoxy-3-methyl-1,4-BQ from beetles fed with unlabeled *Dalotia* diet (control) versus food infused with CD_3-methionine. Spectra were recorded in single-ion mode. A strong $[M+3]^+$ enrichment indicates that the methoxy group is derived from dietary Met. B-D: Expression of *CoQ5* (J; Likelihood-Ratio test BQ-cell versus control: LR-test statistic = 5.37; q = 0.0374), *CoQ6* (K; Likelihood-Ratio test BQ-cell versus control: LR-test statistic = 9.27; q = 0.0082) and *meos* (L; Likelihood-Ratio test BQ-cell versus control: LR-test statistic = 33.74; q < 0.0001) in BQ cells and control tissue based on SMARTseq. The horizontal middle bars represents the median, while the vertical bars show the minimum and maximum. E, F: Representative GC traces of gland compounds from wild-type (*GFP* RNAi; (E) and *meos* RNAi-injected (F) beetles. Silencing of *meos* leads to loss of 2-methoxy-3-methyl-1,4-BQ only, while both other BQs are still present. G: Amounts of 1,4-BQ, 2-methyl-1,4-BQ and 2-methoxy-3-methyl-1,4,-BQ extracted from *GFP* RNAi (control) and MeOS *RNAi*

(legend continued on next page)

beetles. Significance was assessed using Mann-Whitney-U tests (** = $p \leq 0.01$). The horizontal middle bars represents the median, while the vertical bars show the minimum and maximum. H: Expression of *meos* across all cell classes in the 10x data shows tissue specific expression in the sensory neurons. Note that BQ-cells are not captured by 10x. I: MeOS contains a mitochondrial inner membrane (GO:0005743) signal peptide sequence. J: Proposed biosynthesis pathway of *Dalotia*'s BQs versus the mitochondrial ubiquinone pathway. Unique BQ pathway enzymes are in green; unique ubiquinone pathway enzymes are in orange; enzymes shared by the two pathways are in blue.