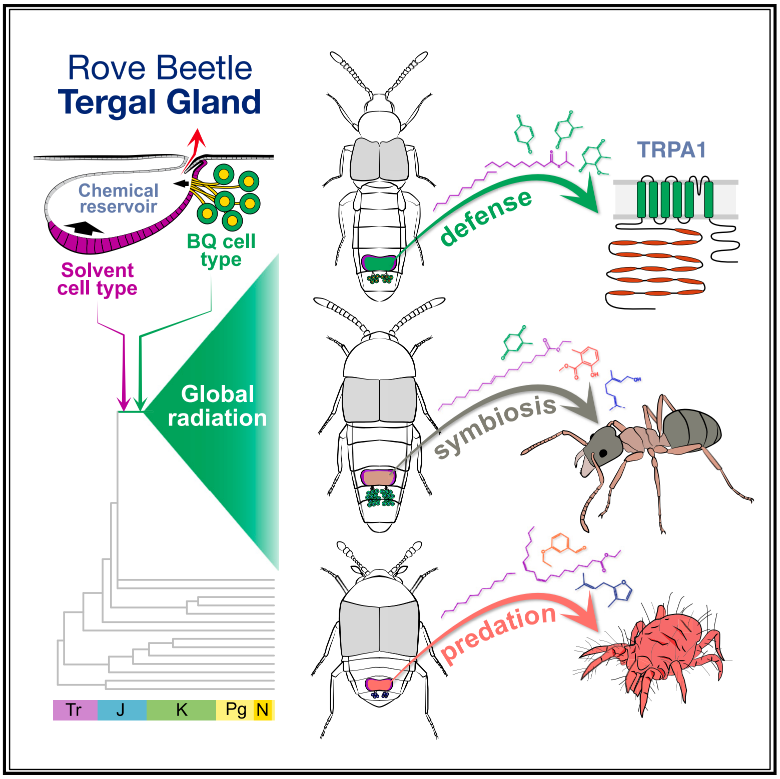
Cell

Article

The genomic and cellular basis of biosynthetic innovation in rove beetles

Graphical abstract



Authors

Sheila A. Kitchen, Thomas H. Naragon, Adrian Brückner, ..., Steven R. Davis, Matthew L. Aardema, Joseph Parker

Correspondence

joep@caltech.edu

In brief

A chemical defense gland is a putative catalyst in the diversification of rove beetles—Metazoa's biggest family. Genomic and cell type-transcriptomic insights retrace the evolution of expression programs encoding cellular mechanisms for defensive compound synthesis and uncover biochemical novelties facilitating ecological specialization.

Highlights

- Two novel cell types form the rove beetle tergal gland, a key evolutionary innovation
- Cellular mechanisms of biosynthesis revealed, encoded by ancient expression programs
- Reprogramming biosynthesis yielded new compounds underlying ecological specialization
- The tergal gland exemplifies cell type innovation driving macroevolutionary success

Kitchen et al., 2024, Cell *187*, 3563–3584
July 11, 2024 © 2024 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.cell.2024.05.012

Cell

CellPress

Article

The genomic and cellular basis of biosynthetic innovation in rove beetles

Sheila A. Kitchen,[1,7] Thomas H. Naragon,[2] Adrian Brückner,[1] Mark S. Ladinsky,[1] Sofia A. Quinodoz,[1] Jean M. Badroos,[2] Joani W. Viliunas,[1] Yuriko Kishi,[1] Julian M. Wagner,[1] David R. Miller,[1] Mina Yousefelahiyeh,[1] Igor A. Antoshechkin,[1] K. Taro Eldredge,[3] Stacy Pirro,[4] Mitchell Guttman,[1] Steven R. Davis,[5] Matthew L. Aardema,[6] and Joseph Parker[1,8,*]

[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]Museum of Zoology, University of Michigan, Ann Arbor, MI 48109, USA
[4]Iridian Genomes, 613 Quaint Acres Dr., Silver Spring, MD 20904, USA
[5]Division of Invertebrate Zoology, American Museum of Natural History, New York, NY 10024, USA
[6]Department of Biology, Montclair State University, Montclair, NJ 07043, USA
[7]Present address: Department of Marine Biology, Texas A&M University, Galveston, TX 77554, USA
[8]Lead contact
*Correspondence: joep@caltech.edu

https://doi.org/10.1016/j.cell.2024.05.012

SUMMARY

How evolution at the cellular level potentiates macroevolutionary change is central to understanding biological diversification. The >66,000 rove beetle species (Staphylinidae) form the largest metazoan family. Combining genomic and cell type transcriptomic insights spanning the largest clade, Aleocharinae, we retrace evolution of two cell types comprising a defensive gland—a putative catalyst behind staphylinid megadiversity. We identify molecular evolutionary steps leading to benzoquinone production by one cell type via a mechanism convergent with plant toxin release systems, and synthesis by the second cell type of a solvent that weaponizes the total secretion. This cooperative system has been conserved since the Early Cretaceous as Aleocharinae radiated into tens of thousands of lineages. Reprogramming each cell type yielded biochemical novelties enabling ecological specialization—most dramatically in symbionts that infiltrate social insect colonies via host-manipulating secretions. Our findings uncover cell type evolutionary processes underlying the origin and evolvability of a beetle chemical innovation.

INTRODUCTION

Exceptional radiations are a recurring pattern across the Tree of Life.[1] Pinpointing ancient genomic and cellular changes that proved to be innovations for the clades that habor them is a major challenge in evolutionary biology.[2] The ~400,000 described beetle species (Coleoptera)[3,4] are an archetype of diversification that has long motivated biologists to consider the causes of species richness.[5–9] The putative beetle key innovation is the elytron—the hardened forewing that shields the delicate flight wings—a structure that enabled beetles to diversify in myriad niches that are inaccessible to other winged insects.[6,10–12] Within Coleoptera, however, diversity is profoundly unbalanced, with ~75% of species belonging to just 10 of 200 extant beetle families. Efforts to explain this biased pattern of diversification have focused primarily on Phytophaga, a megadiverse clade of ~125,000 largely herbivorous species. Phytophagan diversity has been posited to stem from their co-radiation with angiosperms (flowering plants) during the Cretaceous and Cenozoic,[5,13] a phenomenon contingent on key metabolic changes that enabled these beetles to unlock recalcitrant nutrients from plant tissues.[14–18] The catalytic role played by angiosperm herbivory is broadly accepted but leaves open the problem of explaining diversity in the remaining two-thirds of Coleoptera where herbivorous groups comprise only a minority of species.[19] Among the greatest challenges is comprehending the diversity of rove beetles (Staphylinidae)—a clade of 66,464 predominantly predatory species, representing the largest family both in Coleoptera and the whole Metazoa.[20–22]

The extraordinary diversification of rove beetles likely hinged in part on their propensity for chemical innovation, whereby numerous lineages possess abdominal defensive glands with unique, small molecule chemistries.[23,24] These novel structures are thought to have evolved in response to the unusual morphology of staphylinids. Rather than possessing long elytra covering the abdomen, staphylinids typically possess short elytra, exposing a soft, flexible abdomen. This anatomy permits rapid movement through soil and litter but affords little physical protection, fostering widespread evolution of chemical defenses.[20,25] Species richness across the 34 staphylinid subfamilies is strongly skewed, however, with the largest being Aleocharinae—a clade of 16,837 known species,[22] with tens of

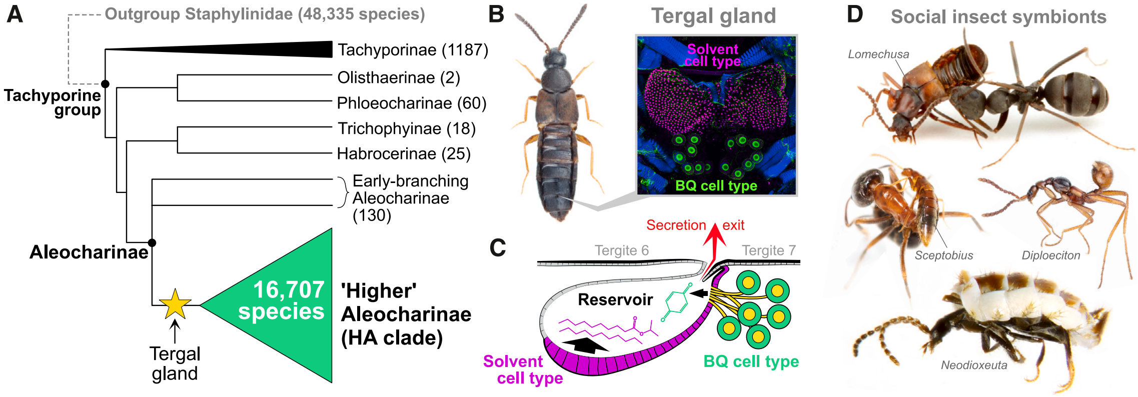


Figure 1. Aleocharine rove beetles

(A) Cladogram of tachyporine-group Staphylinidae[31,32] showing major radiation of Higher Aleocharinae. Numbers in parentheses are described extant species.

(B) Example of a free-living aleocharine (*Atheta* sp.) with confocal image of tergal gland showing position on dorsal body between tergites 6 and 7. The gland comprises two cell types: solvent cells (magenta) and BQ cells (green).

(C) Cartoon of tergal gland showing solvent and BQ cells secreting into common reservoir that ejects between tergites.

(D) Aleocharine symbionts of ants and termites displaying behavioral interactions with hosts (chemical manipulation of host ant by *Lomechusa* and grooming host ant by *Sceptobius*) and symbiotic morphologies (myrmecoid shape of myrmecophile *Diploeciton* and physogastric shape of termitophile *Neodioxeuta*).

thousands more remaining undescribed[26] (Figure 1A). Aleocharines are typically small-bodied (2–6 mm) predators but comprise arguably the most ecologically diverse beetle clade. The group has radiated massively across Earth's temperate and tropical zones, exploiting niches in litter, soil, saproxylic and subcortical microhabitats, fungi, carrion, vascular plants, and environmental extremes in caves, deep soil, intertidal regions, and transiently submerged coral reefs.[27–30] Pervasive ecological and trophic specialization manifests in clades of ectoparasitoids, vertebrate commensals, and social insect symbionts, plus numerous lineages that have shifted to feeding on fungus, dead wood, plants, and pollen.

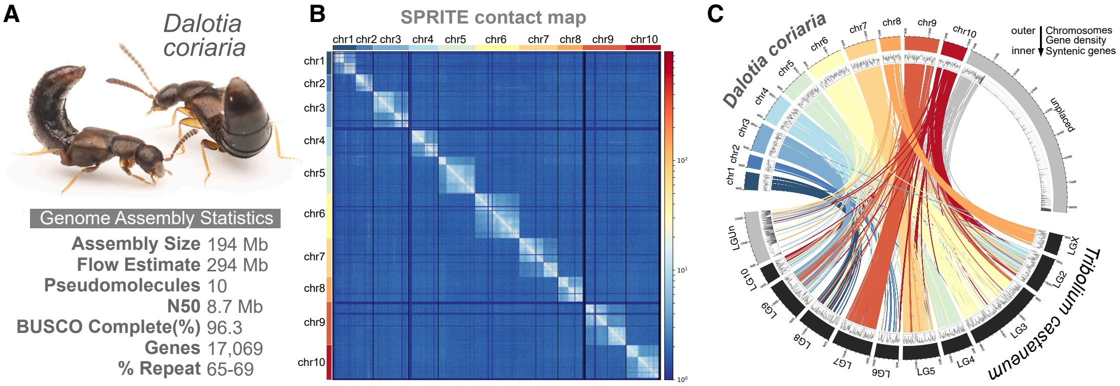
Aleocharinae's unparalleled diversification has been attributed to their defensive "tergal gland"—a dorsal abdominal structure that is targetable at other organisms and exudes a potent, benzoquinone-containing secretion[23,33,34] (Figures 1B and 1C). The gland confers protection against predators such as ants[34–39] and is thought to have enabled aleocharines to radiate explosively in ant-dominated ecosystems worldwide.[40,41] The gland has also been proposed to facilitate infiltration of ant and termite colonies, leading to convergent evolution of symbiotic myrmecophiles and termitophiles across the subfamily[28,40,42–45] (Figure 1D). Tergal gland chemistry has been shown to vary between species, reflecting possible adaptive streamlining to specific niches.[33,39,46–48] The secretion also exhibits antimicrobial properties, potentially aiding colonization of new habitats via pathogen suppression.[34] Crucially, early branching aleocharine lineages and related outgroup subfamilies lack the gland[31,33] and are correspondingly species-poor with limited ecological diversity[40] (Figure 1A). In contrast, the gland is conserved across the 10^4–10^5 so-called "higher Aleocharinae" species, secondarily degenerating only in specialized symbiotic taxa where chemical defense is obsolete.[40,45] The gland is thus a putative key innovation[2,49]—a trait that is correlated with, and likely contributed to, Aleocharinae's remarkable radiation.

Insights into the tergal gland have come from studies of the aleocharine *Dalotia coriaria*, revealing how this structure is composed of two secretory cell types that synergize to produce the defensive secretion.[34] One cell type—the "BQ cells"—converts dietary aromatic amino acids into toxic benzoquinones. These compounds are solids, however, and depend on the second cell type, the "solvent cells," to synthesize fatty acid derivatives into which the benzoquinones dissolve. The resultant cocktail is highly aversive to predators, conferring adaptive value onto this cooperative biosynthetic system.[34] Here, we retrace the evolution of this chemical innovation with a chromosome-level reference genome of *Dalotia coriara*, along with draft assemblies spanning Aleocharinae. By combining comparative genomic and cell-type-specific transcriptomic insights with analyses of enzyme function, gland chemistry, and cellular anatomy, we pinpoint molecular and cellular contingencies that established the tergal gland during early aleocharine evolution. We show that, since its origin, the cell types comprising this structure have exhibited evolutionary stasis at both functional and molecular levels as Aleocharinae radiated into tens of thousands of lineages. Conversely, we find that both cell types have also provided versatile substrates for emergence of biochemical novelties, catalyzing profound niche specialization across this beetle clade. Our findings connect the origin and evolution of new cell types to the macroevolutionary diversification of a major metazoan radiation.

RESULTS

The *Dalotia coriaria* reference genome

To enable broad insights into rove beetle biology, we assembled a high-quality, chromosome-level genome of the laboratory model staphylinid *Dalotia coriaria* (Aleocharinae: Athetini) (Figure 2A). Our approach combined Illumina short paired-end reads (44× coverage) with Oxford Nanopore minION long-reads



Genome Assembly Statistics

Assembly Size	194 Mb
Flow Estimate	294 Mb
Pseudomolecules	10
N50	8.7 Mb
BUSCO Complete(%)	96.3
Genes	17,069
% Repeat	65-69

Figure 2. The *Dalotia* reference genome

(A) Genome assembly statistics of *Dcor* v3.

(B) SPRITE assembled contact map reveals ten chromosomes.

(C) Gene density (density plot, middle band) and synteny (links, inner band) between *D. coriaria* and *T. castaneum* chromosomes or linkage groups (outer band). Inner links colored according to originating *D. coriaria* chromosome. 6.6% of predicted protein-coding genes map to unplaced contigs (gray links).

See also Figures S1–S3.

(54× coverage, $N_{50} = 7{,}933$) for an initial 120 Mb draft assembly, *Dcor* v1 ($N_{50} = 3.97$ Mb, longest scaffold = 12.92 Mb; Figure S1A, Data S1A). Genome-wide heterozygosity remained moderate (k-mer estimate of 1.10%–1.32%) despite seven generations of sibling inbreeding. *Dcor* v1 assembly size approximated that predicted by *k*-mer based tools (139 ± 20 Mb, Figures S1C and S1D), but was less than half the flow cytometry estimate (male 294 ± 11 Mb, female 296 ± 13 Mb). Large discrepancies between *k-mer*- and flow-based genome size estimates have been observed in beetles,[50,51] arising from highly repetitive content.[51] The repeat content of the *Dalotia* genome based on short reads from two separate specimens was 65%–69% (Figures S2A and S2B), composed primarily of a specific 147 bp AT-rich satellite (*Dc-Sat1*) comprising 55% to 61% of the repeatome (Figures S2C and S2D, Data S1B), primarily in intergenic regions (Figure S2E). *Dc-Sat1* is not unique to *Dalotia* but has undergone a species-specific expansion to dominate the repeat landscape (Figures S2C and S2F), consistent with the "library" model of satellite evolution.[52] We found numerous long-reads composed entirely of *Dc-Sat1* arrays and predict that these could form kilobase to megabase-scale, higher-ordered DNA structures (Figures S2G and S2H).[53]

To further extend and orient scaffolds, we generated 262 Bionano optical maps and performed a hybrid assembly with *Dcor* v1. *De novo* assembly of optical maps alone produced a 257 Mb assembly, approaching the flow estimate, but the hybrid assembly with *Dcor* v1 incorporated only 96 of those optical maps, yielding a 122.8 Mb assembly (*Dcor* v2, Figure S1F, Data S1A). We were able to map 883 10 kb or longer minION reads to 124 unincorporated optical maps (74%), suggesting shared repeat structures in long-reads and optical maps that may not be captured in the hybrid assembly (Figure S2G). We uniquely mapped 95% of short- and long-reads to the *Dcor* v2 assembly, indicating abundant repeats like *Dc-Sat1* are present but collapsed in the assembly. We then produced a chromosome-resolved assembly via Split-Pool Recognition of Interactions by Tag Extension (SPRITE),[54] which yields both intra- and interchromosomal contacts (see "*Dalotia* genome assembly" in STAR Methods). After generating a contact map with 11,674,733 clusters identified by SPRITE, we improved contiguity into 10 pseudomolecules, containing 98.9% of the *Dcor* v2 assembly with a scaffold N_{50} of 12 Mb (*Dcor* v3, Figure 2B and Data S1A). The 10 pseudomolecules (hereafter chromosomes) match *Dalotia*'s chromosome count (Figure S1E) and the karyotype of another aleocharine, *Aleochara*.[50] Lastly, we recovered 72 Mb of unincorporated, repeat-rich contigs by mapping the preliminary assemblies back onto *Dcor* v3. These contigs were combined with *Dcor* v3 for a final assembly of 194 Mb (Data S1A).

Gene content in the *Dcor* v3 assembly is near-complete with 96.3% complete/1.3% partial orthologs from the BUSCO arthropod gene set (n = 1,013 genes)[55] (Figure S3A). We predicted 17,069 protein coding genes using transcriptome data spanning life stages and tissue types, predicted gene models from the beetles *Tribolium castaneum* (Tenebrionidae) and *Nicrophorus vespilloides* (Staphylinidae: Silphinae), and *ab initio* tools (see STAR Methods) (Data S1C). 93.4% of the protein-coding genes were found along the 10 chromosomes (Figure 2C). Despite their >250-million-year divergence, gene synteny remains high between *Dalotia* and *Tribolium* (Figure 2C), with 878 syntenic blocks that contain 3–10 shared genes per block. Chr 8 is the probable X chromosome based on significant female-biased expression (χ^2 false discovery rate adjusted $p < 0.001$) and 12.9%, 18.7%, and 8.6% protein conservation with *Tribolium* and the rove beetles *Ocypus olens* and *Philonthus cognatus* (subfamily Staphylininae), respectively (Figure 2C, S3B, and S3D). Chr 1 also had significant female-biased expression (χ^2 false discovery rate-adjusted $p < 0.001$) (Figure S3B). Excessive sex-biased expression from Chr 1 could stem from prior fusion between the ancestral beetle X and Chr 1, resulting in feminization of Chr 1 prior to subsequent fission.[56] Chr 2 is the likely Y chromosome based on significant male-biased expression (Figures S3C and S3D; χ^2 false discovery rate adjusted p = 0.004) but shares little gene content with the *P. cognatus* putative Y (0.2%) (Figure S3B).

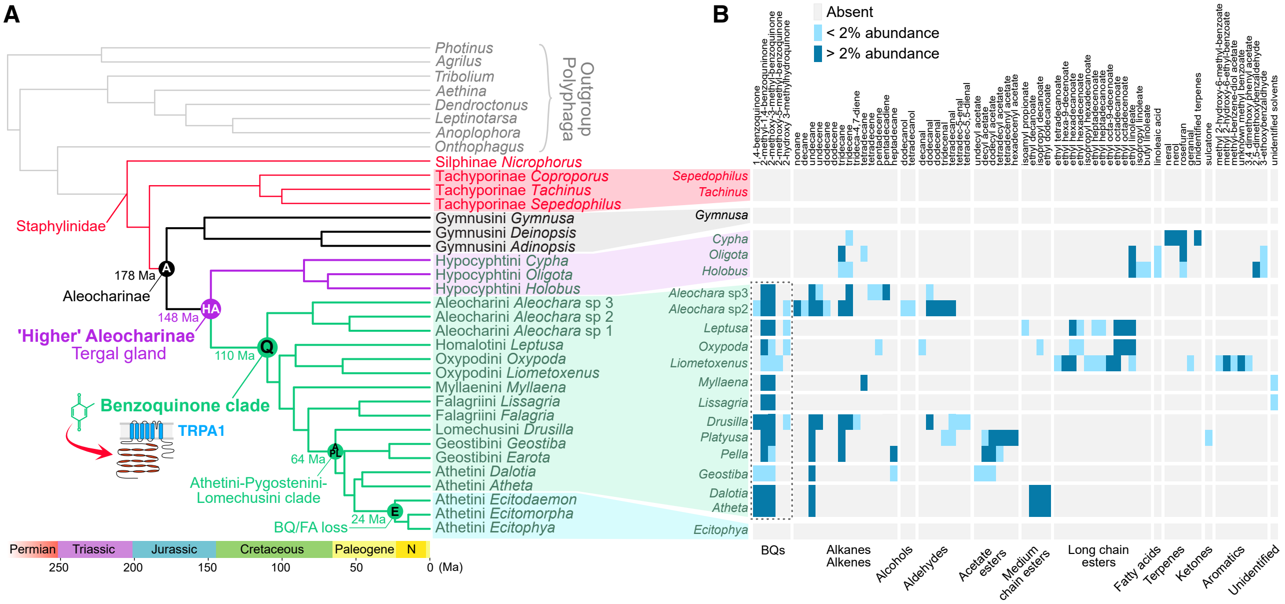


Figure 3. Chemical innovation across Aleocharinae

(A) Dated ML phylogenomic tree inferred from 1,520 orthologs, with key nodes and ages indicated. All nodes received maximal bootstrap support (see also Data S3).

(B) Heatmap of major and minor compounds from aleocharine tergal glands. Dashed box indicates deep conservation of benzoquinones across most aleocharines.

See also Figure S4.

Phylogenomic relationships in Aleocharinae

To explore genome evolution in Aleocharinae, we generated short-read genomic data for a further 24 ingroup and outgroup species, using the nearly complete *Dcor* v1 assembly to guide genome assembly and inform gene predictions (Figure S1B). Taxon sampling was targeted to illuminate traits that arose during early aleocharine evolution, principally the tergal gland. We assembled three genomes from the earliest-diverging, glandless tribe Gymnusini.[26,31–33,57,58] Multiple genomes spanning major gland-bearing higher aleocharine lineages were incorporated, including putative early branching tribes: Hypocyphtini, Aleocharini, and Oxypodini.[31,44,57,58] Taxa from Mylaenini, Falagriini, Homalotini, Geostibini, Lomechusini, and Athetini (to which *Dalotia* belongs) were also included. Among these were genomes of four myrmecophiles to illuminate evolutionary changes in chemistry associated with symbiosis. Three belong to the "*Ecitochara* group" of Athetini (formerly the tribe Ecitocharini)—neotropical ant-mimicking (myrmecoid) symbionts associated with *Eciton* army ants, in which the tergal gland has degenerated.[59] The fourth is *Liometoxenus newtonarum* (Oxypodini), a myrmecophile of *Liometopum* ants from southern California.[60] Outgroup genomes were included from the subfamily Tachyporinae, allied to Aleocharinae within the "Tachyporine-group" of Staphylinidae.[31] Average genome completeness of the new assemblies was 92.6% (range: 54.7%–99.5%) (Figure S3A, Data S1C). Previously published genomes of nine other beetles of high genome completeness were also included, spanning the coleopteran suborder Polyphaga (to which Staphylinidae belongs).

Using 1,520 orthologous protein-coding loci, we inferred a phylogenomic tree of these species, estimating node ages with fossil calibrations within and outside Aleocharinae (Figure 3A, Data S1D and S3A–S3C). Our topology is strongly supported at all nodes (Data S3A and S3B) and broadly congruent with prior phylogenetic studies.[44,57,58] We recovered a monophyletic Aleocharinae, sister to the tachyporines, with a crown-group origin in the Early Jurassic (178 Ma [mega-annum]; 95% highest posterior density [HPD]: 209–150 Ma) (Figure 3A and Data S3C). Within Aleocharinae, glandless Gymnusini are sister to a monophyletic, gland-bearing higher Aleocharinae (clade "HA").[26,31,58,61] We infer that the tergal gland originated close to the Jurassic-Cretaceous boundary, with the HA crown-group dating to 148 Ma (95% HPD: 176–123 Ma). Consistent with previous studies, Hypocyphtini emerge as the earliest-branching HA lineage,[31,57,58,62] with Aleocharini the subsequent HA lineage to diverge. Inside the HA, the homalotine *Leptusa* was recovered as sister to the two oxypodine taxa (*Oxypoda* and *Liometoxenus*), while taxa belonging to the megadiverse "Athetini-Pygostenini-Lomechusini" ("APL") clade[63,64] are recovered as monophyletic, including the tribe Geostibini. We infer an early Paleocene origin of the APL (64 Ma; 95% HPD: 77–53 Ma)—younger than previously estimated[44] (Figure 3A and Data S3C). The APL numbers ~8,600 extant described species and includes the greatest number of myrmecophile and termitophile lineages. Its Cenozoic origin implies an exceptional rate of cladogenesis, with recurrent transitions to social insect symbiosis during a window when modern ants and termites proliferated.[41,65–67] Within the APL, the myrmecophilous *Ecitochara* group is sister to the athetines *Dalotia* and *Atheta*, congruent with earlier studies of athetine relationships[63,64] (Figure 3A).

Chemical evolution in Aleocharinae

We extracted tergal gland secretions from taxa spanning the tree and used gas chromatography-mass spectrometry (GC-MS) to characterize the chemical composition (Figure 3B). The "classical" aleocharine tergal gland secretion employs benzoquinones (BQs) as toxic irritants. Benzoquinones bind TRPA1 channels,[68] activating nociceptive neurons to induce pain. Benzoquinones are solid compounds, however, and are therefore dissolved in a fatty acid (FA)-derived fraction composed of alkanes, alkenes, aliphatic esters, aldehydes, or a combination thereof. The FA-derived solvent unlocks the benzoquinones' potency, creating a noxious secretion.[34] Consistent with previous studies,[33,34] we find such a "BQ/FA cocktail" in most HA taxa (Figures 3A, 3B, and S4). Lineages producing this secretion comprise a vast clade within the HA, herein named the "Q clade" (quinone-producing). The most recent common ancestor (MRCA) of the Q clade existed ∼110 Ma (95% HPD: 132–90 Ma; Figure 3A and Data S3C); *Aleochara* (Aleocharini) represents the earliest-branching Q clade lineage (Figures 3A, 3B, and S4). After the BQ/FA cocktail originated, relative chemical stasis occured almost throughout subsequent cladogenesis, particularly within the benzoquinone fraction, where a small number of variants of 1,4-benzoquinone are conserved across most Q clade taxa (some species also secrete traces of the benzoquinones' hydroquinone precursors) (Figure 3B, dashed box; Figure S4).

FA-derived solvents are similarly conserved, but the precise compounds vary substantially across the Q clade. We and others have previously shown how subtle changes in chain lengths and molar ratios of FA-derivatives strongly influence the secretion's viscosity, wetting ability, and efficacy as a benzoquinone solvent.[34,69] Different lineages have thus modified the physicochemical properties of their secretions. For example, production of medium-chain, acetate-, and some long-chain esters has evolved independently within the Q clade (Figures 3A, 3B, and S4). Esters have been shown to increase the wetting properties of defensive secretions.[70] Low-level production in some taxa is consistent with esters being surfactants rather than the principal solvent.[34,70] In *Dalotia*, esters were also found to be critical for microbial suppression.[34] Esters may therefore represent a recent adaptive addition in certain lineages. Moreover, esters have superseded alkanes as the primary solvent in the oxypodines *Oxypoda* and *Liometoxenus* and the homalotine *Leptusa*—potentially via a single secondary loss or reduction of alkanes in their MRCA (Figures 3A, 3B, and S4). Curiously, both alkanes and esters have been lost in the falagriine *Lissagria*, consistent with earlier chemical data from Falagriini.[33] Presently unidentified compounds may be solvents in falagriines. Evolvability of the secretion is further underscored by taxa scattered across the tree incorporating novel compound classes, including ketones, terpenes, and other aromatics (Figure 3B). Tergal gland chemistry therefore appears to be reprogrammable during evolution, potentially facilitating ecological specialization. The tergal gland can also become dispensable: members of the *Ecitochara* group have secondarily lost benzoquinones and any solvents (Figures 3B and S4), consistent with gland degeneration in these myrmecophiles.[59]

Stasis in gland cell type evolution

We asked how changes at the genomic, pathway, and cell type levels underlie evolution of tergal gland chemistry. Two secretory cell types comprise the gland: "BQ cells" that manufacture benzoquinones and "solvent cells" that produce FA derivatives into which the benzoquinones dissolve (Figures 1B and 1C). Previously, we generated BQ and solvent cell type-specific transcriptomes from *Dalotia coriaria*, enabling us to elucidate biosynthetic pathways for *Dalotia*'s BQ/FA cocktail.[34] To gain insight into the origins and functional evolution of BQ and solvent cells, we sought to retrace their evolution across the HA clade. *Dalotia*'s secretion contains three benzoquinones; these are dissolved in a large volume of a medium-chain alkane, undecane, along with three aliphatic esters: ethyl decanoate, isopropyl decanoate, and ethyl dodecanoate (Figure 4A, upper trace). The earliest-branching lineage producing a comparable BQ/FA is *Aleochara* (Aleocharini), demarcating the Q clade that encompasses the HA minus the tribe Hypocyphtini (Figure 3A). Although *Aleochara* diverged from *Dalotia* in the early Cretaceous, 110 Ma (Figure 3A and Data S3C), *Aleochara* species nevertheless produce two or all three of the same benzoquinones as *Dalotia* (Figure 3B). Similarly, these benzoquinones are dissolved in alkanes, predominantly undecane and tridecane; some *Aleochara* secretions additionally contain aldehydes (alkane precursors) and alkenes. Unlike *Dalotia*, *Aleochara* secretions do not contain esters (Figures 3B and S4).[33,71]

We assembled a draft genome of a southern Californian *Aleochara* (sp. 3 in Figure 3), the secretion from which shares with *Dalotia* two benzoquinones (2-methyl-1,4-BQ and 2-methoxy-3-methyl-1,4-BQ) and undecane (Figure 4A, lower trace). We dissected replicates of BQ and solvent cells from this *Aleochara* and assembled cell type-specific transcriptomes via SMART-seq, creating a dataset directly comparable to that obtained from homologous cell types in *Dalotia* (Data S1E, S4A, and S4B). Microdissection resulted in 3–7 BQ cells, ∼1,000 solvent cells, or ∼1,000 control cells from tergite 6 per replicate (Figure 4B; see STAR Methods). Due to differences in sequencing library preparation, we assessed the impact of potential sources of technical variation on *Dalotia* and *Aleochara* datasets individually. In both, variation was highest among individual samples, followed by differences between tergal gland cell types, with only a minor or no contribution attributable to technical variation (Data S1F and S2). To compare expression between species, we restricted our analysis to 9,314 orthologs shared between the two beetles and transformed read counts using an empirical Bayes method to remove effects attributable to species.[72] Gene expression evolution between *Dalotia* and *Aleochara* BQ, solvent, and other abdominal cell types was explored via principal component analysis (PCA) on replicate cell type-specific transcriptomes. Strikingly, each tergal gland cell type of *Dalotia* clustered with the homologous cell type from *Aleochara*, with strong separation of BQ and solvent cells both from each other and from control tissue (Figure 4C). Hence, despite the ∼110 Ma separation between *Aleochara* and *Dalotia*, their BQ and solvent cells each differentially express common gene sets, potentially underlying conservation of the BQ/FA cocktail across the Q clade.

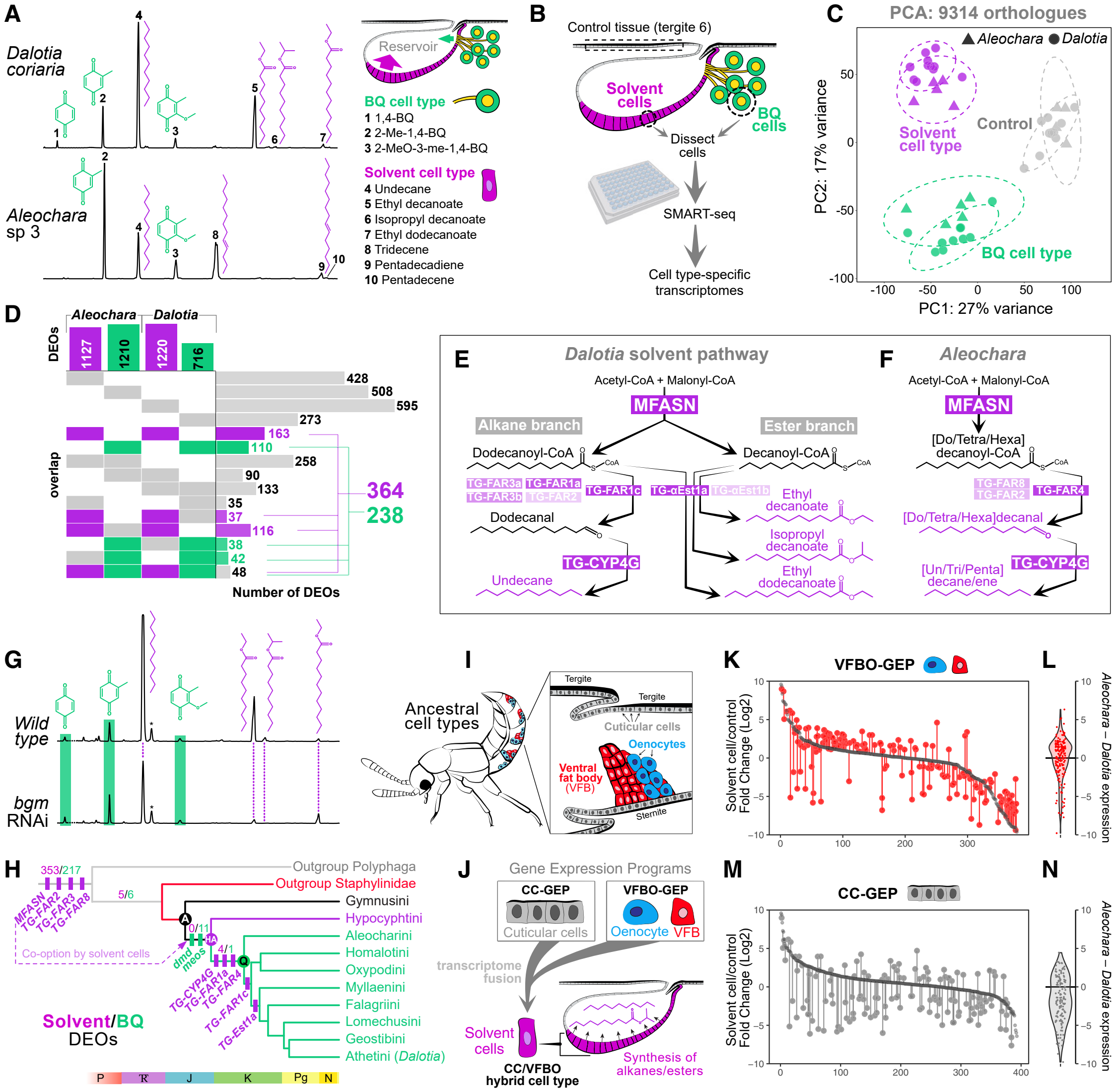


Figure 4. Deep conservation of tergal gland gene repertoire in the Q clade

(A) GC traces of *Dalotia* and *Aleochara* compounds and their cell type of origin.

(B) Scheme for cell-type-specific transcriptomes.

(C) PCA of all expressed orthologs (n = 9,314) in *Dalotia* and *Aleochara* solvent cells, BQ cells, and control tissue (tergite 6).

(D) UpSet plot showing shared DEOs for each cell type by species and cell type.

(E and F) Solvent pathways in *Dalotia* and *Aleochara*, with cases of paralog co-expression in solvent cells. Transparency of purple boxes equates to maximum $\log_2$ fold-change above control tissue for paralogs.

(G) Example GC traces from wild-type *Dalotia* (top trace, n = 14) and *bgm*-silenced animals (n = 42).

(H) Time-calibrated tree showing origins of key enzymes.

(I and J) Schematic of abdominal cell types with gene expression programs (GEPs) for ventral fat body/oenocytes and cuticle cells (I), hybridization of which created the solvent cell type (J).

(K and M) *Aleochara* solvent cell expression (red or gray) relative to *Dalotia* solvent cell expression (black) for orthologs of the highest Z score ranked genes in VFBO-GEP and CC-GEP.

(L and N) Violin plots showing difference in *Aleochara* from *Dalotia* solvent cell expression for genes within each GEP.

See also Figure S5.

A conserved solvent cell expression program

We examined transcriptomic similarity between *Dalotia* and *Aleochara* tergal gland cell types and identified 364 diffentially expressed orthologs (DEOs) in solvent cells of both species and 238 DEOs shared by their BQ cells (Figure 4D). These DEOs define deeply conserved “core” gene expression programs within each cell type. We asked whether these programs might encode ancient biosynthetic toolkits within the Q clade and discovered that tergal gland cells of *Dalotia* and *Aleochara* express homologous pathways for defensive compound biosynthesis. In *Dalotia* solvent cells, alkane and ester synthesis derives from a bifurcating fatty acid pathway in which a fatty acid synthase, Master FASN (MFASN), produces C10 and C12 fatty acid precursors (Figure 4E). In one downstream pathway branch, the C12 fatty acid is reduced to an aldehyde by a fatty acyl-CoA reductase (Tergal Gland FAR1c; TG-FAR1c, formerly "TG-FAR"[34]); the aldehyde is then decarbonylated by a 4G-class cytochrome P450 (TG-CYP4G), yielding undecane. In a parallel branch, the C10 fatty acid is esterified by a carboxylesterase of the α-esterase family (TG-αEst1a; formerly "TG-αEst"[34]), resulting in the two C10 esters (Figure 4E). TG-αEst1a also esterifies traces of the C12 fatty acid, making ethyl dodecanoate (Figure 4E).

Key components of this pathway are deeply conserved in the Q clade. In *Aleochara*, MFASN is again the sole fatty acid synthase expressed in solvent cells (Figures 4F and S5A, Data S1E and S1G); likewise, the decarbonylase *TG-CYP4G* comprises part of the core solvent expression program (Figures 4F and S5B, Data S1E and S1G). Multiple other core components have predicted roles in solvent biosynthesis, and the core program is significantly enriched in biological processes related to fatty acid synthesis and modification (Data S1H). One previously uncharacterized step in solvent production is the activation of fatty acids produced by MFASN by addition of CoA.[73] Among core transcripts, we identified a very long-chain-fatty-acid-CoA synthase (LC-FACS), orthologous to the *Drosophila* gene *bubblegum* (*bgm*) (Figure S5C). Silencing *bgm* in *Dalotia* with RNAi caused significant reduction in undecane (41% of GFP control, Wilcoxon signed-rank with Bonferroni p adjusted = 0.005) and near-complete loss of ethyl decanoate (12% of GFP control, p adjusted < 0.001; Figures 4G and S5D). Bgm is thus at least partially responsible for activation of fatty acid precursors of defensive alkanes and esters in Q clade aleocharines.

Beyond the core program of orthologous loci, functionally equivalent paralogs can be identified in *Aleochara* and *Dalotia*. In total, 27 *FAR* copies are encoded in the *Dalotia* genome and 21 in *Aleochara* (Data S3D). *Dalotia* solvent cells express five *FAR* paralogs (Figure 4E), one of which, *TG-FAR1c*, accounts for virtually all undecane synthesis.[34] In every *Aleochara* genome we surveyed, however, a *TG-FAR1c* ortholog was absent (Data S3D). Instead, *Aleochara* solvent cells express three *FAR* paralogs—*TG-FAR2*, *4*, and *8*, one or more of which likely performs the equivalent step in alkane synthesis (Figure 4F and Data S3D). The *FAR* family undergoes extensive gene birth-and-death in insects[74]; weak expression of *TG-FAR2* in *Dalotia* solvent cells may be a vestige of its earlier involvement in alkane production prior to the birth of *TG-FAR1c* (Figure 4H; Data S3D). One key difference between the two beetles' pathways is the ester branch, present only in *Dalotia* (Figures 4E and 4F). *Dalotia*'s ester production is mediated by *TG-αEst1a* that lacks an apparent ortholog in *Aleochara* (Data S3E). Indeed, no α-esterases or other carboxylesterases are expressed in *Aleochara*'s solvent cells (Data S1E and S3E). Appending an ester branch was therefore a more recent innovation in solvent pathway evolution, enabled by the birth of *TG-αEst1a*.

Solvent cell evolution through ancient transcriptome hybridization

Notably, 353 of 364 loci in the solvent cell core expression program are co-opted genes with orthologs across Polyphaga (Figure 4H). Strong predominance of co-option may stem from how solvent cells are thought to have originated. They are a secretory cell type but form part of the beetle's exoskeleton. Using single-cell RNA-seq of *Dalotia*'s abdominal cell types, we previously showed that solvent cells are a hybrid of two gene expression programs—one that defines cuticular identity (the "cuticular cell" gene expression program [CC-GEP]) and another that defines two ancient metabolic cell types: ventral fat body cells and oenocytes that produce cuticular hydrocarbon pheromones (ventral fat body/oenocyte-GEP [VFBO-GEP])[34] (Figure 4I). VFBO-GEP is strongly enriched for loci involved in fatty acid synthesis and modification,[34] implying that VFBO-GEP co-option into cuticular cells endowed the latter with solvent-producing capacity (Figure 4J). We examined whether VFBO-GEP and CC-GEP are conserved in *Aleochara*. We first ranked *Dalotia* loci according to their Z score computed previously by Brückner et al.[34] within both VFBO-GEP and CC-GEP and then compared expression of each locus in *Dalotia* solvent cells to that of its ortholog in *Aleochara* solvent cells. Strikingly, VFBO-GEP loci are also differentially expressed in *Aleochara* solvent cells, with relative expression of these orthologs strongly correlated between *Aleochara* and *Dalotia* (n = 288 orthologs; Spearman rho = 0.61, $p < 0.001$; Figures 4K and 4L; Data S4C). Conversely, conservation of CC-GEP in solvent cells is weaker: fewer *Aleochara* orthologs show comparable expression in *Dalotia* solvent cells (n = 303 orthologs; Spearman rho = 0.15, p = 0.009; Figures 4M and 4N, Data S4C). These findings imply that formation of solvent cells, via recruitment of VFBO-GEP into cuticle cells, was an ancient event, pre-dating the Q clade MRCA. Subsequent conservation of VFBO-GEP in solvent cells occurred despite marked divergence in the cuticular program.

Evolution of benzoquinone production and the BQ cell type

Akin to solvent cells, we find evidence of deep molecular conservation within the BQ cell type. In *Dalotia*, benzoquinones derive from aromatic amino acids such as tyrosine[34] (Figures 5A and 5B). These are converted to 4-hydroxybenzoic acid (4-HB), which is modified in the BQ cell mitochondrion via sequentially acting ubiquinone/coenzyme Q pathway enzymes. The resultant hydroquinones are then thought to be secreted into the BQ cell lumen where they undergo oxidation by a secreted laccase, Decommissioned (Dmd), converting them into the final, toxic benzoquinones (Figures 5A and 5B). Critical components of this pathway are conserved in *Aleochara*. As in *Dalotia*, *Aleochara* possess a single *dmd* ortholog that is strongly upregulated in BQ cells (Figure 5B and Data S4A and S4B). We synthesized and purified *Aleochara* Dmd and found it efficiently converts 2-methyl-1,4-hydroquinone to the corresponding benzoquinone (Figure 5C), implying hydroquinone oxidation by Dmd is an ancient, terminal step in Q clade benzoquinone biosynthesis (Figure 5B). Upstream mitochondrial steps also appear conserved. Like most Q clade taxa, *Dalotia* and *Aleochara* produce 2-methoxy-3-methyl-1,4-BQ (Figure 4A). In *Dalotia*, the methoxy

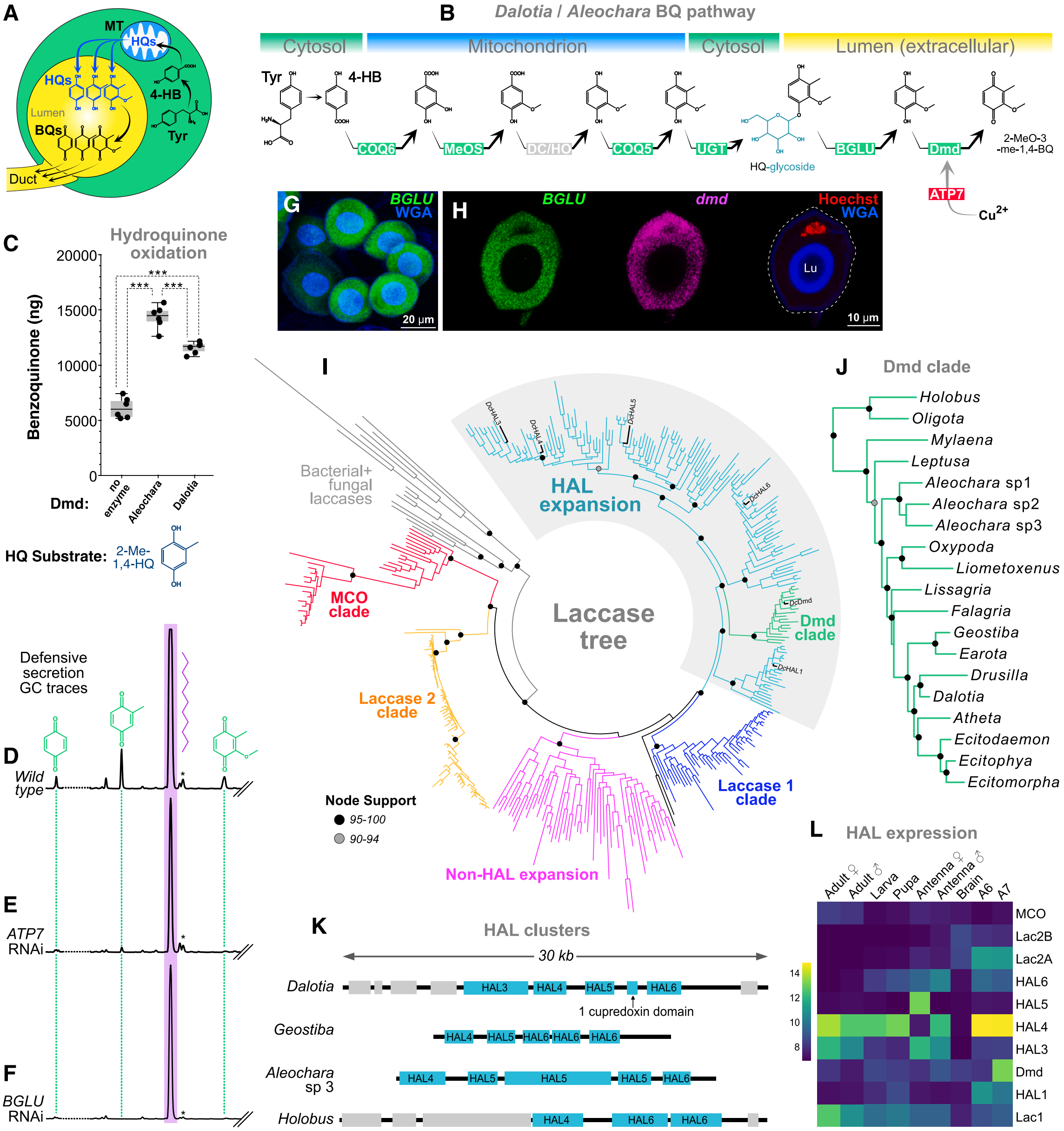


Figure 5. Evolution of BQ chemistry

(A) Cartoon of BQ cell showing benzoquinone synthesis from tyrosine (Tyr).

(B) Benzoquinone pathway in Q clade Aleocharinae, showing cellular locations of enzymatic steps.

(C) *In vitro* conversion of 2-methyl-1,4-hydroquinone to benzoquinone by purified *Aleochara* or *Dalotia* Dmd. Asterisks denote $p < 0.0001$ in Tukey post-hoc tests.

(D–F) Tergal gland GC traces from wild-type *Dalotia* (D), *ATP7*-silenced animals (E), and *BGLU*-silenced animals (F). Dotted line indicates hexane contamination peaks (removed for clarity). Asterisks denote peaks of dimethyl-BQ spiked in as positive control.

(G and H) *BGLU* expression in *Dalotia* BQ cells. Green, *BGLU* HCR; blue, WGA; in (G), magenta is *dmd* HCR and red is Hoechst-labeled nucleus. Lu, lumen.

(I) ML tree of laccase gene family showing higher Aleocharine laccase (HAL) expansion in light blue; *Dalotia* HAL paralogs are indicated (substitution model LG + R10 with 1,000 bootstrap replicates; support for larger clades is shown by the circle color: black = 95%–100% and gray 94%–90%).

(J) Expanded Dmd clade from (I) reveals conservation across HA taxa. Node support values < 90% are not displayed.

(K) Genomic HAL clusters of selected aleocharine taxa.

(L) Expression heatmap of *Dalotia* laccases, including HALs, from RNA-seq data obtained from tissues, life stages, and sexes.

See also Figure S6.

group is added by a mitochondrial enzyme, Methoxyless (MeOS)—an aleocharine-specific duplicate of COQ3 that adds a methoxy group to ubiquinone.[34] We recovered the single *meos* ortholog in *Aleochara* within the BQ cell type's core expression program (Figures 5B and S6A, Data S4A and S4B).

Overall, the core BQ program is enriched in biological processes related to mitochondrial metabolism and metal ion transport (Data S1H). Other core transcripts represent newly discovered components with putative functions in benzoquinone production. Laccases related to Dmd are known to depend on elevated import of Cu^{2+}—a process in mammals mediated by ATPase transporters ATP7A/B and the copper chaperone ATX1.[75,76] Conspicuously, both the single-copy aleocharine homologs of mammalian *ATP7A/B* and *ATX1* comprise part of the BQ cell type's core program (Figure S6B, Data S4A and S4B). Silencing *ATP7* in *Dalotia* strongly diminished levels of the highest abundance benzoquinone, 2-methyl-1,4-BQ (Wilcoxon signed-rank with Bonferroni p adjusted = 0.0267, Figures 5D, 5E, and S6C). Elevated Cu^{2+} in BQ cells is likely essential for Dmd activity, providing the cofactor for this metalloenzyme (Figure 5B).

Upstream of Dmd, the mechanisms of intracellular trafficking of hydroquinone precursors were previously unknown. Despite the widespread use of benzoquinones in arthropod chemical defenses,[77] it has been unclear how cells are safeguarded from these cytotoxic compounds.[34,78] In plants, small molecule toxins are often conjugated to sugars, creating relatively harmless glycosides that are hydrophilic, facilitating intracellular storage and transport.[79,80] Upon herbivory, the glycoside is commonly released from cells to undergo cleavage by a β-glucosidase that removes the sugar moiety, activating the toxin.[81] An analogous mechanism was previously hypothesized for benzoquinone regulation in insects.[82] Remarkably, within the BQ cell type's core expression program is a predicted β-glucosidase (BGLU) (Data S4A and S4B), expression of which was confirmed by *in situ* hybridization chain reaction (HCR) (Figures 5G and 5H). Strikingly, silencing this *BGLU* in *Dalotia* led to near-complete elimination of all benzoquinones from the secretion (Wilcoxon signed-rank test with Bonferroni p adjusted < 0.001 for each compound; Figures 5D, 5F, and S6D). Glycosides may thus indeed be the form in which the BQ cells produce hydroquinones, prior to β-glucosidase-mediated cleavage. The *Dalotia BGLU* encodes a secreted protein, implying that hydroquinone glycosides are secreted into the BQ cell lumen, prior to cleavage by BGLU and oxidation by Dmd (Figure 5B). Among the most strongly upregulated core transcripts in both *Dalotia* and *Aleochara* BQ cells is a UDP-glycosyltransferase (UGT): an enzyme with a classical role in conjugating toxins to glucose or related sugars.[83,84] UGT is thus a candidate enzyme for producing hydroquinone glycosides (Figure 5B). These results uncover a mechanism of benzoquinone regulation that has convergently evolved with small-molecule chemical defense mechanisms in plants. Further characterization of these enzymes *in vitro* awaits identification of their specific glycoside substrates *in vivo*.

As in solvent cells, the BQ cell core expression program is composed predominantly of ancient, co-opted genes, with 217/238 loci having orthologs across Polyphaga (Figure 4H). Twelve loci, however, are aleocharine-specific novelties, which arose in HA or Q clade stem lineages and may have potentiated benzoquinone evolution. One of these is the *COQ3* paralog *methoxyless* (*meos*), which originated along the HA stem and experienced positive selection (CodeML LRT = 19.63, p < 0.001; Figures 4H and S6A). *COQ3* is a single-copy gene in most eukaryotes.[85] However, it has repeatedly duplicated in both aleocharines and tachyporines (Figure S6A), yielding four copies in *Dalotia* including *meos* (Figure 5B).[34] Most notably, *dmd* itself is found exclusively in HA genomes (Figure 5J). We retraced *dmd*'s origin and found it emerges within a major, monophyletic expansion of laccase enzymes in HA genomes. This "higher Aleocharine laccase" (HAL) clade encompasses 6 *Dalotia* paralogs but up to 15 in other species (Figure 5I and Data S3F). Significant episodic selection occured on almost all branches leading to the major splits in the HAL expansion, suggesting neofunctionalization of these duplicates (aBSREL select branch test, p < 0.05, Data S3F). HAL copies can be dispersed within the genome, but many sit tandemly in a single genomic cluster (Figure 5K). In *Dalotia*, each HAL copy is expressed in a different tissue pattern, developmental stage, or sex, implying distinct functions (Figure 5L). Curiously, an independent laccase expansion exists in the glandless Gymnusini and outgroup tachyporines (Figure 5I). This "non-HAL" expansion must predate Aleocharinae but has been lost in higher aleocharines and replaced with the HAL expansion. Genomes of most insects encode only three conserved laccases (Figure 5I), including laccase 2 that functions in pigmenting and sclerotizing the cuticle.[86,87] Laccases in general are known for oxidizing phenolic compounds,[88] and we speculate that HALs may have enabled aleocharines to better detoxify soil-, plant-, or fungal-derived phenolics (to which these beetles must be routinely exposed). A byproduct of the HAL expansion was birth of a duplicate—Dmd—which would ultimately become neofunctionalized for benzoquinone synthesis.

Gland conservation and divergence in the earliest-branching HA lineage

Our findings uncover an ancient gland toolkit in the Q clade that has been preserved as these beetles radiated over ~110 Ma. Yet, the tergal gland predates the Q clade: this structure is a synapomorphy of the HA, encompassing the Q clade and a further, early-branching lineage: the small tribe Hypocyphtini (Figure 3A).[31,57,58,62] Hypocyphtini may provide critical insights into tergal gland evolution but remain unexplored beyond confirming their possession of a solvent reservoir (supporting their systematic placement in HA[31,62]). Hypocyphtines are enigmatic in being mite predators, some providing biocontrol of pest mite species.[89,90] This specialized biology contrasts with the generalist predatory lifestyle thought to be ancestral in Aleocharinae. Morphologically, hypocyphtines are also divergent, with a minute, compact body and short abdomen (Figures 6A–6C and S6E). Due to Hypocyphtini's key phylogenetic position, we assembled draft genomes and profiled secretions of three genera covering the tribe's diversity: *Cypha*, *Oligota*, and *Holobus* (Figure 3A).

All three beetles produce an alkane/alkene: tridecane/tridecene (Figures 6A–6C). Further, two genera produce a long-chain fatty acid, linoleic acid, and ester derivatives thereof, revealing

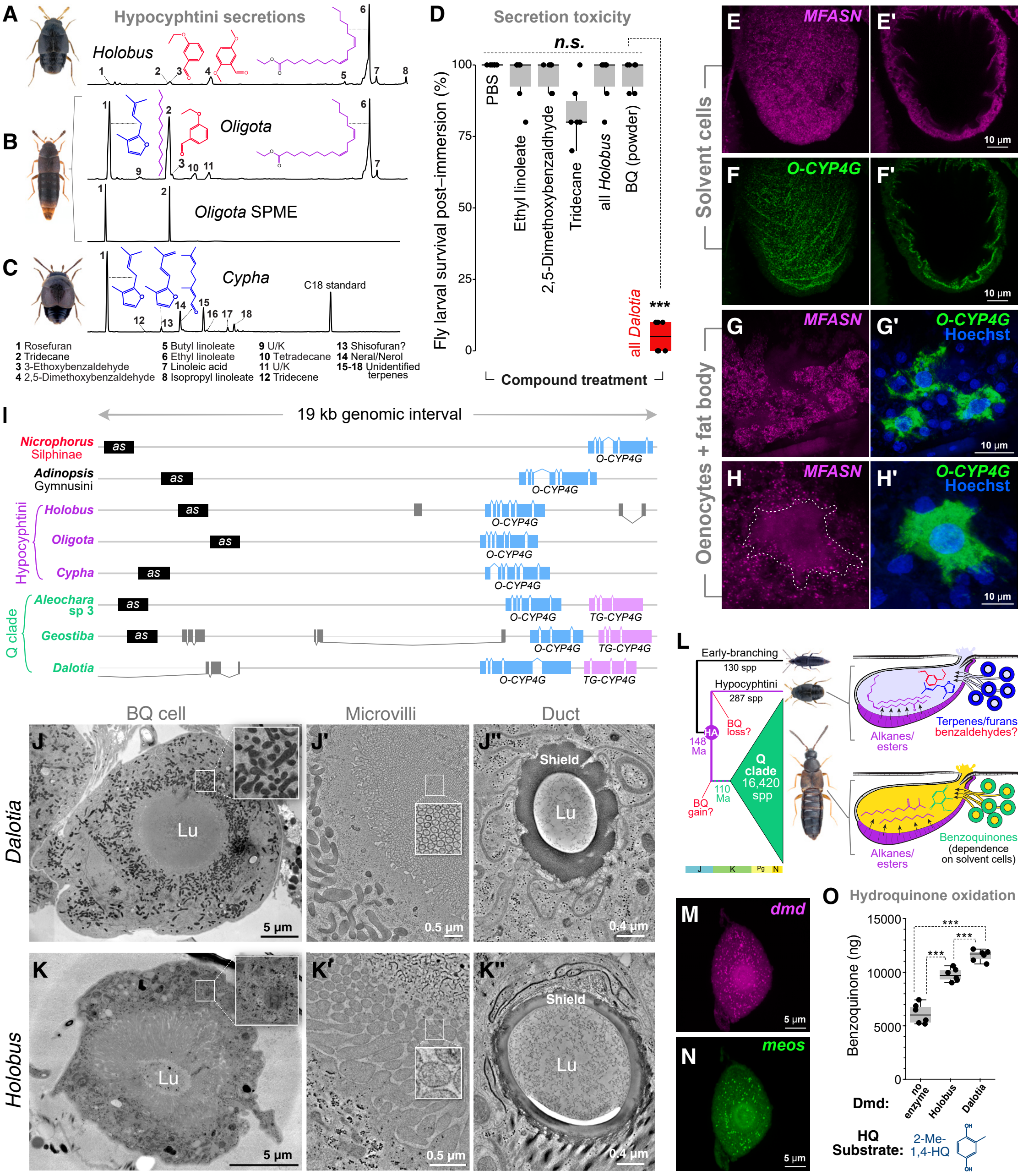


Figure 6. Glandular biology of the earliest-branching HA lineage

(A–C) GC traces of hypocyphtine glandular compounds: *Holobus* (A), *Oligota* (B, B′), and *Cypha* (C). (B′) shows headspace volatiles from 20 *Oligota* beetles detected via SPME.

(D) *Drosophila* larval survival following immersion in synthetic hypocyphtine or *Dalotia* secretions. Outcome of Tukey post-hoc test between treatments is shown (n.s., not significant; ***$p < 0.0001$ in all individual comparisons between “all *Dalotia*” gland compounds and the other treatments).

(E and F) HCR of *MFASN* (E, E′, magenta) and *O-CYP4G* (F, F′, green) in *Oligota* solvent reservoir (E, F: labeling within plane of solvent cell epithelium; E′, F′: cross section through reservoir).

(G and H) HCR of *MFASN* (magenta) and *O-CYP4G* (green) in *Oligota* fat body and oenocytes (blue, Hoechst-stained nuclei).

(I) Synteny reveals origin of *TG-CYP4G* in Q clade (*Aleochara* sp. 3, *Geostiba*, and *Dalotia*) via duplication of *O-CYP4G*, present as a single copy in Hypocyphtini (*Holobus*, *Oligota*, and *Cypha*), the glandless gymnusine *Adinopsis*, and outgroup silphine *Nicrophorus*. The upstream gene *asense* (*as*) is a conserved syntenic feature of all species except *Dalotia*. For further details of synteny, see Data S5A.

(legend continued on next page)

conservation of FA-derived solvents across the HA (Figure 3B). Examining a species of *Oligota*, we HCR labeled the solvent pathway fatty acid synthase *MFASN*, revealing expression in both solvent cells and abdominal fat body (Figures 6E, 6G, and 6H), mirroring the pattern in *Dalotia* (Figure S5E).[34] *MFASN* was thus co-opted into solvent cells in the HA stem (rather than the Q clade stem) and its function has been conserved there as the HA radiated throughout the Cretaceous and Cenozoic (Figure 4H). In contrast, the decarbonylase *TG-CYP4G* is absent from all hypocyphtine genomes (Figures 6I and S5B and Data S5A). *TG-CYP4G* is thus a Q clade novelty (Figure 4H). *TG-CYP4G* is a duplicate of an ancient cytochrome P450, *Oenocyte-CYP4G* (*O-CYP4G*), that is conserved across Coleoptera (and Insecta; Figure S5B). *O-CYP4G* functions in oenocytes to decarbonylate very long chain aldehydes, yielding cuticular hydrocarbon pheromones (CHCs).[91,92] Remarkably, in hypocyphtines, it is *O-CYP4G* that is expressed in solvent cells (Figure 6F), in addition to oenocytes (Figures 6G and 6H), implying *O-CYP4G* was first co-opted into solvent cells prior to duplicating. Following duplication, *TG-CYP4G* experienced episodic selection: 12 codons show signatures of positive selection and three others show relaxed selection within the Q clade (aBSREL $\omega_2 = 8.32$, LRT = 18.74, $p < 0.001$; CodeML LRT = 27.79, $p < 0.001$) (Figure S5B). Simultaneously, *O-CYP4G* experienced positive selection post-duplication (Figure S5B). We infer that *O-CYP4G* was co-opted into solvent cells in the HA stem and functioned pleiotropically in both CHC and defensive alkane synthesis—a situation preserved in hypocyphtines. Subsequently, the gene duplicated in the Q clade stem, yielding oenocyte and solvent cell copies. Freed from pleiotropic constraint, both copies underwent adaptive evolution. The tandem syntenic arrangement of *O-CYP4G* and *TG-CYP4G* has been conserved across the Q clade (Figure 6I and Data S5A).

The most remarkable feature of hypocyphtine secretions is the absence of benzoquinones. Instead, all three beetles secrete a furan, rosefuran (Figures 6A–6C); further, *Cypha* produces monoterpenes (from which rosefuran is likely derived). Additionally, both *Holobus* and *Oligota* produce benzaldehydes—compounds unseen in other aleocharines. We relate these chemical novelties to Hypocyphtini's acariphagous biology. Rosefuran is a mite sex pheromone,[93] the monoterpene neral is a mite attractant or alarm pheromone,[94,95] and benzaldehyde pheromones are widespread in mites.[94,96–98] Consequently, we propose that hypocyphtines possess gland chemistries specialized for mite predation. Chemical defense seems unlikely: the furan and terpenes lack pronounced toxicity or irritant properties, nor do the benzaldehydes, which we tested by immersing *Drosophila* larvae in 2,5-dimethoxybenzaldehyde (produced by *Holobus*). This compound caused no reduction in survival when applied either alone or mixed with the specific alkane and ester that *Holobus* produces (Figure 6D, 2,5-dimethoxybenzaldehyde compared to PBS control Tukey post-hoc $p = 0.99$; *Holobus* gland cocktail compared to PBS control Tukey post-hoc $p = 0.99$). In contrast, potent lethality results from immersion in synthetic *Dalotia* tergal gland secretion (Figure 6D, *Dalotia* gland cocktail compared to PBS control Tukey post-hoc $p < 0.0001$). Reduced abdominal mobility of hypocyphtines likely precludes them from directly smearing secretions on other organisms—the mode of deployment in many aleocharines.[35,37] Sampling headspace volatiles above *Oligota* beetles, we detected strong secretion of rosefuran and tridecane but no linoleic acid derivatives, which appear not to be volatilized (Figure 6B, lower trace). We hypothesize that volatilized rosefuran may provide chemical mimicry, or act as a chemical lure during mite predation.

We examined cellular ultrastructure within the tergal gland using electron tomography, confirming that hypocyphtines possess BQ cells like those of *Dalotia*. *Dalotia* BQ cells are large (~30 μm diameter) spherical acini, with a lumen formed by involution of the apical membrane (Figure 6J). Dense microvilli extend into the lumen, presumably secreting hydroquinone glycosides together with BGLU and Dmd for conversion to benzoquinones (Figure 6J′). Connected to each BQ cell is a long, convoluted duct, enveloping a lumen with a thick, protective shield for channeling benzoquinones into the gland reservoir (Figure 6J″). BQ cells of the hypocyphtine *Holobus* are smaller (~15 μm diameter) but share this overall anatomy (Figure 6K). Both solvent and BQ cells are thus HA synapomorphies, dating to the MRCA of HA at the Jurassic-Cretaceous boundary (Figure 3A). HA and hypocyphtine BQ cells nevertheless differ in key aspects, most strikingly in their mitochondrial content. *Dalotia* BQ cells are extremely rich in mitochondria, consistent with a high demand for hydroquinone synthesis by these organelles (Figure 6J, inset). Conversely, *Holobus* BQ cells have scarce mitochondria (Figure 6K, inset), consistent with them not synthesizing benzoquinones but instead mite pheromones. Other ultrastructural differences may correspond to a reduced need for protection from cytotoxicity: the lumenal microvilli are thicker and less densely organized (Figure 6K′), and the duct lumen is wider and less heavily shielded (Figure 6K″).

Due to the minute size of hypocyphtines (Figure S6E), we have been unable to dissect their tergal glands for cell-type transcriptomics. The pathways these cells express remain unknown. However, Hypocyphtini's lack of benzoquinones raises a fundamental question about the BQ cell type's ancestral function. Hypocyphtines may embody a transitional stage in tergal gland evolution prior to benzoquinones originating in the Q clade stem (Figure 6L). Alternatively, benzoquinones may have arisen in the HA stem and been secondarily lost in hypocyphtines (Figure 6L). Curiously, two marker genes of benzoquinone synthesis—*dmd* and *meos*—are present in hypocyphtine genomes (Figures 5J and S6A, Data S3F). Moreover, both are expressed

---

(J and K) TEM of *Dalotia* (top) and *Holobus* (bottom) BQ cells. Lu, lumen. Insets in (J) and (K) show differing mitochondrial densities between the two species (electron-dense structures). (J′) and (K′) show differing microvillar organization and density within BQ cell lumens. J″ and K″ show differing shield thickness within internal lumen (Lu) of ducts.

(L) Topology of deepest divergences in Aleocharinae. Alternative scenarios posit benzoquinones were gained in Q clade or lost in Hypocyphtini.

(M and N) HCR of *dmd* (M, magenta) and *meos* (N, green) in *Oligota* BQ cells.

(O) *In vitro* conversion of 2-methyl-1,4-hydroquinone to benzoquinone by purified *Holobus* or *Dalotia* Dmd. Asterisks denote $p < 0.0001$ in Tukey post-hoc tests.

in BQ cells of *Oligota* (Figures 6M and 6N). Bulk RNA-seq further revealed elevated expression of *ATP7* and *BGLU* in gland-bearing abdominal segment 7 of *Holobus* (Data S1I). We synthesized and purified *Holobus* Dmd and found it produced significant 2-methyl-1,4-BQ when provided with hydroquinone (albeit less efficiently than *Dalotia* Dmd; Tukey post-hoc tests $p < 0.0001$) (Figure 6O). These findings might be interpreted as evidence of an intermediate evolutionary stage in the BQ cell type's core expression program: key components are present, but not yet assembled into a functional pathway. Conversely, these components may equally represent "molecular spandrels"[99,100]—ghosts of functions past, providing evidence of the cell type's prior role in benzoquinone production.[99,100] That *Holobus* Dmd retains activity may imply a new function within BQ cells or elsewhere in the beetle. We consider hypocyphtine chemistry to be highly specialized for mite predation and unlikely to represent the primitive condition in HA. Hence we posit hypocyphtines have lost benzoquinones. Overall, hypocyphtine BQ and solvent core loci have been under relatively relaxed selection compared to Q clade orthologs (Data S5B), consistent with hypocyphtine chemistry being derived. Consequently, we propose that cooperation between solvent and BQ cells, yielding the BQ/FA cocktail,[34] may have been present in the MRCA of the entire HA clade, 148 Ma.

Evolvability of tergal gland cell types under symbiosis

Hypocyphtine secretions reveal how the tergal gland has provided an evolutionary substrate for specialized chemical interactions. Aleocharine chemical innovation is well known for being taken to the extreme in symbiotic lineages specialized for life within social insect colonies. Symbionts have been demonstrated to use secretions to confuse, pacify, or appease workers, or to elicit beetle adoption into the nest.[38–40,42,43,47,48,101–106] Several taxa have been hypothesized to have repurposed the tergal gland to produce host-manipulating secretions, implying biosynthetic reprogramming of BQ and/or solvent cell types.[43,46–48] Pursuing this phenomenon, we discovered dramatic modification of tergal gland chemistry in the myrmecophile *Liometoxenus*—a genus described recently for which no prior chemical, behavioral, or genomic data existed[60] (Figure 7A). *Liometoxenus* inhabits colonies of *Liometopum* ants in Southern California. We observe the beetles executing a remarkable behavioral interaction with host workers where *Liometoxenus* secretes a volatile cocktail that acts at a distance to intoxicate ants, impairing their locomotion and attenuating aggression toward the beetle (Video S1). This manipulation enables *Liometoxenus* to prey upon workers.

We profiled *Liometoxenus* tergal gland chemistry and found a complex cocktail containing 18 compounds spanning multiple classes: long- and medium-chain aliphatic esters (both saturated and unsaturated); benzoquinones identical to those of free-living species; a long series of aromatic esters; and a terpene, geranial (Figure 7B). To our knowledge, this secretion is the most diverse chemical mixture produced by a single rove beetle species. We assembled a draft genome of *Liometoxenus newtonarum* and created cell type-specific transcriptomes for both BQ and solvent cells via SMART-Seq (Data S1J). Using 8,641 orthologs shared between *Liometoxenus*, *Dalotia*, and *Aleochara*, we performed PCA on these three species, which again clustered homologous cell types with each other along PC1 and PC2, showing deep conservation of their core transcriptomes (Figure S7A). Along PC3, however, we observed strong separation of *Liometoxenus* BQ cells from those of *Dalotia* and *Aleochara* (Figure S7B). Investigating the BQ cell transcriptome further, we found evidence of dramatic pathway evolution in *Liometoxenus* (Figure 7C). First, we identified an entire monoterpene synthesis pathway in BQ cells, presumably leading to geranial. Enzymes for every step from mevalonate-5-phosphate to geraniol pyrophosphate are present (Figures 7C–7E). We cannot identify with certainty the terminal geranial synthase (GES), but BQ cells express *Liometoxenus*-specific duplicates of both FPPS and GGPS—enzymes that in all known terpene-producing insects have convergently duplicated and neofunctionalized into terpene synthases[107–109] (Figures S7C and S7D). We posit a parallel scenario in *Liometoxenus*.

Synthesis of benzoquinones appears to be identical to other aleocharines, with expression of all known pathway components conserved in *Liometoxenus* BQ cells (Figure 7C). However, the metabolic precursor of benzoquinones—tyrosine—has become strongly biased toward synthesis of new compounds. Feeding *Liometoxenus* adults Tyr-$^{13}C_6$ led to strong ^{13}C incorporation into the benzoquinones, as in *Dalotia*[34] (Figure 7F), but also into the aromatic esters that dominate the secretion (their molecular weights increasing by +6; Figure 7G). Like the benzoquinones, these compounds are thus not sequestered from the diet, nor are their benzene rings synthesized *de novo*, but we cannot presently infer their biosynthetic origin. Unlike *Dalotia* and *Aleochara* solvents, headspace sampling revealed that the long-chain esters of *Liometoxenus* are non-volatile (Figure S7E), creating a solvent from which the remaining compounds volatilize to influence ant behavior from a distance. The solvent precursors are likely palmitic and stearic acid (C16 and C18; Figure 7D)—among the commonest insect fatty acids, deriving from lipogenesis in the fat body[110,111]; however, additional synthesis within solvent cells is likely given high expression of enzymes driving the fatty acid elongation cycle (Figure 7C), perhaps accounting for ester chain length variation (Figure 7B). Unlike *Dalotia*'s ester pathway, *Liometoxenus* does not employ an α-esterase; instead, carboxylesterases of alternative families function in solvent cells and may carry out esterification (Figures 7C and 7D, Data S3E). Most esters are present in both saturated and unsaturated forms (Figure 7B), the latter presumably due to expression of the canonical metazoan stearic/palmitic acid desaturase, SCD (stearoyl-CoA desaturase) (Figure 7D).

Liometoxenus uses its secretion to manipulate worker behavior but appears not to engage in complex social interactions with ants. In the most highly integrated symbionts, however, noxious defenses are less critical as the beetles evolve social behaviors and chemical mimicry that assimilate them into host societies.[45] In several such taxa, the tergal gland has evolutionarily degenerated.[59,112,113] In our phylogenomic sampling, we included members of one such clade—the *Ecitochara* group (*Ecitophya*, *Ecitomorpha*, and *Ecitodaemon*).[59] These beetles are myrmecoid ant mimics (Figure 7H), which are accepted into nomadic colonies of *Eciton* army ants.[59,114] As a first

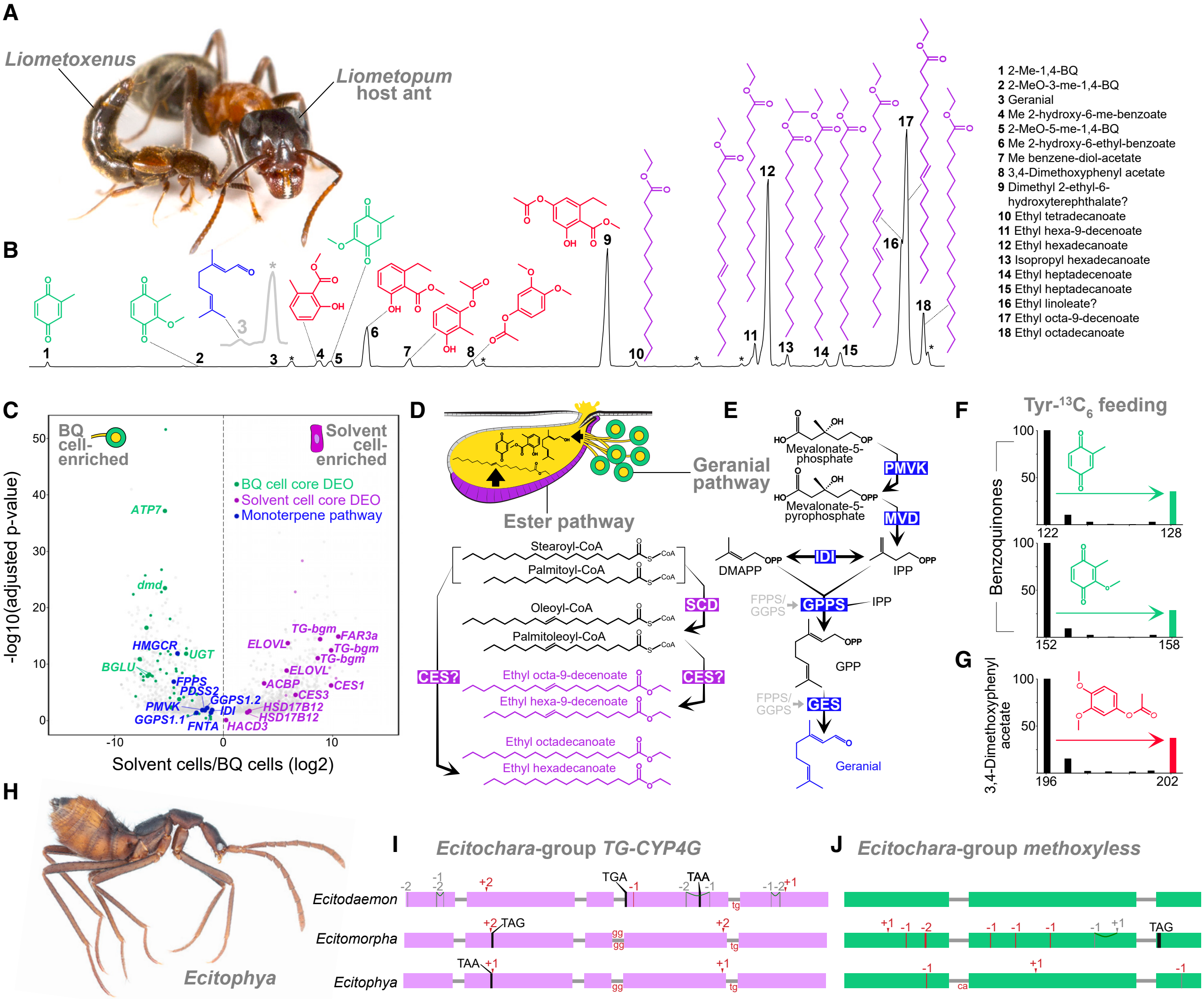


Figure 7. Tergal gland evolution in myrmecophiles

(A) *Liometoxenus newtonarum* with *Liometopum occidentale* host (photo by David Miller).

(B) GC trace of *Liometoxenus* gland compounds. Magnification of geranial peak (compound 3) in gray. Asterisks: contaminants.

(C) Volcano plot of *Liometoxenus* solvent cells (positive $\log_2$ fold-change) and BQ cells (negative $\log_2$ fold-change). DEOs encoding key enzymes are colored (solvent cell, purple; BQ cell, green) along with novel enzymes including inferred monoterpene pathway (blue). *IDI*: *isopentenyl-diphosphate delta-isomerase 1*; *FPPS*: *farnesyl pyrophosphate synthase*; *HMG-CoA*: *3-hydroxy-3-methylglutaryl-coenzyme A reductase*; *PMVK*: *phosphomevalonate kinase*; *FNTA*: *farnesyl-transferase/geranylgeranyltransferase type-1 subunit alpha*; *GGPS1_1/GGPS1_2*: *geranylgeranyl pyrophosphate synthase*; *PDSS2*: *decaprenyl-diphosphate synthase subunit 2*; *SCD5.2*: *stearoyl-CoA desaturase 5*; *CES1*: *Carboxylesterase 4A*; *ACBP*: *Acyl-CoA-binding protein homolog*; *CES3*: *type-B carboxylesterase*.

(D) Cartoon showing hypothesized pathway for *Liometoxenus* aliphatic esters. SCD, Acyl-CoA Delta(11) desaturase. "CES" denotes hypothesized function of either or both carboxylesterase 4A (CES1) or type-B carboxylesterase (CES3) in solvent cells.

(E) Inferred terpene pathway leading to geranial.

(F and G) Mass spectra of molecular ion regions of compounds from *Liometoxenus* fed with dead ants infused with $^{13}C_6$-Tyr. Spectra were recorded in single-ion mode. 2-methyl-1,4-BQ (MW 122) and 2-methoxy-3-methyl-1,4,-BQ (MW = 152) exhibit strong $[M+6]^+$ enrichment (green bars) (F), as does 2-hydroxy-6-methyl-benzoate (red bar).

(H) *Ecitophya simulans* beetle.

(I and J) *TG-CYP4G* and *methoxyless* gene models from *Ecitochara*-group species showing inactivating mutations. Negative/positive numbers are frameshift base pair deletions/insertions against the reference genome (*Dalotia*). Premature stop codons are shown; splice junction mutations are shown at intron-exon boundaries.

See also Figure S7.

glimpse into how such an integrated lifestyle impacts genomic evolution, we analyzed the BQ and solvent cell core expression programs in these beetles. Consistent with most core loci having been co-opted into the tergal gland, 35%–38% of 554 BQ and solvent cell core loci remain present (intact or partially intact) in the genomes of these myrmecophiles. However, of those intact, 63 and 41 loci were under significant relaxed selection or intensified (positive or purifying) selection, respectively, relative to the other higher aleocharines, suggesting possible loss or divergence of function following gland degeneration (RELAX analysis FDR < 0.05; Data S5B). Loci under relaxed selection include *ATP7*, *dmd*, *stearoyl-CoA desaturase TG-SCD*, and *fatty acyl-CoA reductase FAR2*.

While many core loci were missing or partially missing (21%–33%) due to fragmentation of genome assemblies, we found clear evidence of pseudogenization and gene loss in 13, 10, and 12 core biosynthetic genes from *Ecitodaemon*, *Ecitomorpha*, and *Ecitophya*, respectively (Data S1K). Multiple inactivating mutations, including frameshifts and premature stop codons, have accumulated in both the solvent cell decarbonylase *TG-CYP4G* and the benzoquinone-modifying enzyme *methoxyless* (Figures 7I and 7J). Such a pattern of gene inactivation is consistent with removal of purifying selection following degeneration of the now-obsolete tergal gland. Specific inactivating mutations are often not shared by all three taxa, with only *TG-CYP4G* of *Ecitophya* and *Ecitomorpha* sharing a subset of changes (Figure 7I). Moreover, *Ecitodaemon* still possesses an intact *methoxyless* (Figure 7J). Given that the three genera share an MRCA ∼24 Ma (95% HPD: 33–15 Ma; Figure 3A), in which the gland had presumably already degenerated, these idiosyncratic patterns of gene-inactivating mutations imply a surprisingly slow rate of coding sequence decay in these myrmecophiles. All three species also possess an apparently intact *dmd* ortholog (Figure 5J and Data S3F), which when expressed *in vitro* converted hydroquinones to benzoquinones (Figure S7F; Tukey post-hoc tests $p < 0.001$). We posit that Dmd plays an alternative role in these myrmecophiles.

DISCUSSION

The radiation of Metazoa's largest family, Staphylinidae, has been coupled to pervasive biochemical innovation, precipitated by convergent evolution of abdominal exocrine glands. Here we examined the evolution of one such structure—the aleocharine tergal gland. We uncovered evolutionary changes at the genome, pathway, and cell type levels that underlie the gland's assembly in early aleocharines, its deep functional conservation as the beetles radiated globally, and its potential for evolvability via biosynthetic repurposing.[33,39,46–48] Our findings underscore how new organismal properties can derive from *de novo* evolution of cell types, with ramifications at the macroevolutionary scale.

Assembly and stasis of gene expression programs for defensive chemistry

We inferred that the solvent and BQ cells comprising the tergal gland arose early in aleocharine evolution, along the HA stem. These cell types and their secretions have been broadly conserved across the HA, numbering tens of thousands of lineages that began diversifying in the Early Cretaceous. Macroevolutionary patterns at the genomic, transcriptomic, cell type, and chemical levels imply that long-term stabilizing selection on defensive chemistry has occurred almost clade-wide across the HA. At the cellular level, this is reflected in conservation of the BQ and solvent cells and their cooperative interaction, manifesting in relative evolutionary stasis of core expression programs conferring each cell type's biosynthetic function. Our findings emphasize how modular gene expression programs are fundamental units on which natural selection can operate to sustain or build novel cell type and organ functions.[115] Each core expression program comprises a majority of phylogenetically ancient, co-opted loci, with a small handful of recent paralogs encoding key enzymes. Expansions along the HA stem of laccases and COQ3 duplicates were decisive genomic contingencies yielding enzymes for benzoquinone synthesis. Gene co-option in novel cell types has been hypothesized as a source of pleiotropic conflict from which duplication permits escape.[78] Such a scenario may explain the origins of taxon-restricted loci in BQ and solvent cells. TG-CYP4G, originating via duplication of O-CYP4G, provides a clear case of co-option prior to duplication, followed by adaptive evolution of both copies (embodying the "escape from adaptive conflict" model of duplicate gene evolution[116,117]).

Ecological specialization through cell type evolvability

Broad conservation of aleocharine defensive chemistry has not precluded dramatic evolutionary innovations in biosynthesis. Fatty acid derivatives produced by solvent cells can vary extensively, with predicted effects on the secretion's physicochemical properties. Such streamlining may enable production of a functional secretion despite microclimatic differences, or permit alternative modes of gland deployment (e.g., directly smearing the total secretion onto targets as in *Dalotia*, versus volatilizing the non-solvent fraction from a distance, as in *Liometoxenus* and *Oligota*). The BQ cells have also proven highly modifiable, producing probable mite pheromones in hypocyphtines or the high complexity secretion of *Liometoxenus* that facilitates ant colony infiltration. The BQ cell type's anatomy and its employment of a plant-like system of toxin regulation involving glycoside cleavage, implies a versatile system that may be co-opted for production of other compounds. How tergal gland cell types gain new multi-enzyme pathways presents a conundrum, since a battery of loci must become co-expressed within the same cell simultaneously to create a compound that renders each locus visible to natural selection. The recruitment of gene expression programs into tergal gland cell types by "terminal selector" transcription factors is likely.[78,118] Candidates are the Hox proteins Abdominal A and Abdominal B that are needed for BQ and solvent cell differentiation[119] but also remain active post-differentiation. We speculate that these transcription factors play governing roles in the evolvability of tergal gland chemistry.

A perplexing finding is the transcription of enzymes that we infer functioned ancestrally within tergal gland cell types but no longer apparently influence biosynthesis. These enzymes may perform new roles within these cell types, but their persistent expression may also derive from enhancer pleiotropy: regulatory elements that drive expression both in the tergal gland and in other organismal contexts where their gene products are visible to natural selection. That Dmd orthologs from the hypocyphtine *Holobus* and myrmecophile *Ecitophya* are functionally intact indicate involvement of this laccase elsewhere in metabolism, or possibly in production of benzoquinones in larvae—a life stage that in other aleocharines has been shown to produce a BQ/FA cocktail from a developmentally distinct abdominal gland.[120]

Cell type evolution of a key innovation

The inordinate diversity of beetles is thought to have been contingent on evolution of protective elytra.[6,10–12] Paradoxically, the largest and most ecologically diverse beetle family has partially forsaken this trait, reducing elytron size to expose the abdomen. Staphylinid cladogenesis may, ironically, have hinged on this loss of physical protection, elytron reduction opening a path to an alternative mode of protection in the form of targetable defensive glands. The evolution of novel cell types comprising peripheral structures such as exocrine glands can profoundly modulate the interaction between an organism and its environment.[78] Analogous to the origin of photoreceptors[121] or cnidocytes,[122] the tergal gland may be a more recent example where *de novo* evolution of cell types has enabled a clade to enter many new adaptive zones. Through chemical and antimicrobial defense, the gland has bought aleocharines enemy-free-space[123] to colonize and diversify throughout Earth's terrestrial ecosystems. As a reprogrammable device, the gland has enabled aleocharines to evolve specialized ecological relationships with other species. Such direct connections between the tergal gland and Aleocharinae's numerical and ecological diversity implicate this structure and its two cell types as a key innovation behind one of Coleoptera's most successful radiations.

Limitations of the study

Future efforts to improve the *Dalotia* genome with high-fidelity long reads may enable better resolution of repeats, including the abundant *Dc-Sat1* satellite. While we have demonstrated adaptive evolution of key biosynthetic enzymes, we have not connected these changes to protein function. Currently, *in vivo* studies of gene function are feasible in *Dalotia*; optimization of these methods is needed to explore gene function in other aleocharines.

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
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
    - ○ Rove beetles
    - ○ Sample size and inclusion/exclusion criteria
- METHOD DETAILS
    - ○ DNA extraction and short and long-read sequencing
    - ○ Bionano optical mapping
    - ○ SPRITE
    - ○ Illumina genome assemblies
    - ○ *Dalotia coriaria* genome assembly
    - ○ Repeat identification and masking
    - ○ *Dalotia* gene predictions and annotation
    - ○ Gene predictions of other genome assemblies
    - ○ Phylogenomic tree construction and dating
    - ○ Phylogenetic analyses of select gene families
    - ○ Selection tests and inactivating mutations
    - ○ Gene synteny
    - ○ Gland volatile quantification
    - ○ Ancestral state reconstruction
    - ○ Biochemical tracer experiment in *Liometoxenus*
    - ○ Double-stranded RNA synthesis and knockdown
    - ○ *Drosophila* toxicity bioassay
    - ○ Chromosome squashes
    - ○ Electron microscopy and dual-axis tomography
    - ○ RNA extraction and transcriptome assemblies
    - ○ SMART-seq transcriptome sequencing
    - ○ Differential expression analysis
    - ○ *In situ* hybridization chain reaction
    - ○ *In vitro* measurement of dmd enzymatic activity
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.05.012.

ACKNOWLEDGMENTS

We thank Charlie Barnes, Mike Caterino, Munetoshi Maruyama, Thomas Schmitt, and Christoph von Beeren for beetle specimens; Taku Shimada and Udo Schmidt for *Dalotia* photography and specimen images; Nathan Dalleska (Water and Environment lab, Caltech) for HPLC assistance, and Elizabeth Soehalim for help with SPRITE. We acknowledge support from Caltech's Center for Evolutionary Science. A.B. was a Simons Fellow of the Life Sciences Research Foundation. J.M.W. and J.W.V. are NSF GRFP recipients. This work was funded by grants to J.P. from the NIH (1R34NS118470-01), NSF (2047472 CAREER), along with a Shurl and Kay Curci Foundation grant, Rita Allen Foundation Scholarship, Pew Biomedical Scholarship, and an Alfred P. Sloan Foundation fellowship. Additional funding was provided by Iridian Genomes (IRGEN_RG_2021-1345 Genomic Studies of Eukaryotic Taxa), Caltech's Millard and Muriel Jacobs Genetics and Genomics Laboratory, and an American Museum of Natural History Gerstner Fellowship in Bioinformatics and Computational Biology to M.L.A.

AUTHOR CONTRIBUTIONS

Conceptualization, J.P., M.L.A., and S.A.K.; methodology, J.P., M.L.A., and S.A.K.; investigation, S.A.K., T.H.N., A.B., M.S.L., S.A.Q., J.M.B., J.W.V., Y.K., J.M.W., D.R.M., M.Y., I.A.A., K.T.E., S.P., M.G., S.R.D., M.L.A., and J.P.; formal analysis, S.A.K., T.H.N., A.B., J.M.B., and J.W.V.; data curation, S.A.K.; writing – original draft, J.P. and S.A.K.; writing – review & editing, J.P. and S.A.K.; supervision, J.P.; project administration, J.P.; funding acquisition, J.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 7, 2023
Revised: February 29, 2024
Accepted: May 6, 2024
Published: June 17, 2024

REFERENCES

1. Simpson, G.G. (1944). Tempo and Mode in Evolution (Columbia University Press).
2. Erwin, D.H. (2021). A conceptual framework of evolutionary novelty and innovation. Biol. Rev. *96*, 1–15. https://doi.org/10.1111/brv.12643.
3. Hammond, P. (1992). Species Inventory. In Global Biodiversity, B. Groombridge, ed. (Spinger), pp. 17–39. https://doi.org/10.1007/978-94-011-2282-5_4.
4. Zhang, Z.Q. (2013). Order Coleoptera Linnaeus, 1758. In: Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa *3148*, 203–208.
5. Farrell, B. (1998). "Inordinate Fondness" explained: why are there so many beetles? Science *281*, 555–559. https://doi.org/10.1126/science.281.5376.555.
6. Grimaldi, D.A., and Engel, M.S. (2005). Evolution of the Insects (Cambridge University Press).
7. Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O.S., Wild, R., Hammond, P.M., Ahrens, D., Balke, M., Caterino, M.S., et al. (2007). A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. Science *318*, 1913–1916. https://doi.org/10.1126/science.1146954.
8. Berry, A. (2008). Ardent beetle-hunters: natural history, collecting, and the theory of evolution. In Natural selection and beyond: the intellectual legacy of Alfred Russel Wallace, C.H. Smith and G. Beccaloni, eds. (Oxford University Press), pp. 47–65.
9. McKenna, D.D., Wild, A.L., Kanda, K., Bellamy, C.L., Beutel, R.G., Caterino, M.S., Farnum, C.W., Hawks, D.C., Ivie, M.A., Jameson, M.L., et al. (2015). The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. Syst. Entomol. *40*, 835–880. https://doi.org/10.1111/syen.12132.
10. Crowson, R.A. (1981). The Biology of the Coleoptera (Academic Press).
11. Lawrence, J.F., and Newton, A.F. (1982). Evolution and Classification of Beetles. Annu. Rev. Ecol. Syst. *13*, 261–290. https://doi.org/10.1146/annurev.es.13.110182.001401.
12. Linz, D.M., Hu, A.W., Sitvarin, M.I., and Tomoyasu, Y. (2016). Functional value of elytra under various stresses in the red flour beetle, *Tribolium castaneum*. Sci. Rep. *6*, 34813. https://doi.org/10.1038/srep34813.
13. McKenna, D.D., Sequeira, A.S., Marvaldi, A.E., Farrell, B.D., and Berenbaum, M.R. (2009). Temporal Lags and Overlap in the Diversification of Weevils and Flowering Plants. Proc. Natl. Acad. Sci. USA *106*, 7083–7088.
14. Kirsch, R., Gramzow, L., Theißen, G., Siegfried, B.D., ffrench-Constant, R.H., Heckel, D.G., and Pauchet, Y. (2014). Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. Insect Biochem. Molec. *52*, 33–50. https://doi.org/10.1016/j.ibmb.2014.06.008.
15. McKenna, D.D., Shin, S., Ahrens, D., Balke, M., Beza-Beza, C., Clarke, D.J., Donath, A., Escalona, H.E., Friedrich, F., Letsch, H., et al. (2019). The evolution and genomic basis of beetle diversity. Proc. Natl. Acad. Sci. USA *116*, 24729–24737. https://doi.org/10.1073/pnas.1909655116.
16. Seppey, M., Ioannidis, P., Emerson, B.C., Pitteloud, C., Robinson-Rechavi, M., Roux, J., Escalona, H.E., McKenna, D.D., Misof, B., Shin, S., et al. (2019). Genomic signatures accompanying the dietary shift to phytophagy in polyphagan beetles. Genome Biol. *20*, 98. https://doi.org/10.1186/s13059-019-1704-5.
17. Salem, H., Bauer, E., Kirsch, R., Berasategui, A., Cripps, M., Weiss, B., Koga, R., Fukumori, K., Vogel, H., Fukatsu, T., and Kaltenpoth, M. (2017). Drastic Genome Reduction in an Herbivore's Pectinolytic Symbiont. Cell *171*, 1520–1531.e13. https://doi.org/10.1016/j.cell.2017.10.029.
18. Salem, H., Kirsch, R., Pauchet, Y., Berasategui, A., Fukumori, K., Moriyama, M., Cripps, M., Windsor, D., Fukatsu, T., and Gerardo, N.M. (2020). Symbiont Digestive Range Reflects Host Plant Breadth in Herbivorous Beetles. Curr. Biol. *30*, 2875–2886.e4. https://doi.org/10.1016/j.cub.2020.05.043.
19. Barraclough, T.G., Barclay, M.V., and Vogler, A.P. (1998). Species richness: Does flower power explain beetle-mania? Curr. Biol. *8*, R843–R845. https://doi.org/10.1016/s0960-9822(07)00527-1.
20. Parker, J. (2017). Staphylinids. Curr. Biol. *27*, R49–R51. https://doi.org/10.1016/j.cub.2016.07.050.
21. Betz, O., Irmler, U., and Klimaszewski, J. (2018). Biology of Rove Beetles (Staphylinidae), O. Betz, U. Irmler, and J. Klimaszewski, eds. (Springer).
22. Newton, A. (2022). Staphyliniformia world catalog database. In Catalogue of Life Checklist https://doi.org/10.48580/dfqt-3gk.
23. Dettner, K. (1993). Defensive secretions and exocrine glands in free-living staphylinid beetles—their bearing on phylogeny (Coleoptera: Staphylinidae). Biochem. Systemat. Ecol. *21*, 143–162.
24. Francke, W., and Dettner, K. (2004). Chemical Signalling in Beetles. Top. Curr. Chem. *240*, 85–166. https://doi.org/10.1007/b98316.
25. Hammond, P.M. (1979). Wing-folding mechanisms of beetles, with special reference to investigations of adephagan phylogeny (Coleoptera). In Carabid beetles: their evolution, natural history, and classification, T.L. Erwin, G.E. Ball, and D.R. Whitehead, eds., pp. 113–180.
26. Hammond, P. (1975). The phylogeny of a remarkable new genus and species of gymnusine staphylinid (Coleoptera) from the Auckland Islands. Syst. Entomol. *44*, 153–173. https://doi.org/10.1111/j.1365-3113.1975.tb00010.x.
27. Seevers, C.H., and Herman, L.H. (1978). A generic and tribal revision of the North American Aleocharinae (Coleoptera: Staphylinidae). Fieldiana Zool. *71*, i–289. https://doi.org/10.5962/bhl.title.3136.
28. Yamamoto, S., Maruyama, M., and Parker, J. (2016). Evidence for social parasitism of early insect societies by Cretaceous rove beetles. Nat. Commun. *7*, 13658. https://doi.org/10.1038/ncomms13658.
29. Newton, A.F., Thayer, M.K., Ashe, J.S., and Chandler, D.S. (2000). Staphylinidae Latreille 1802. In American Beetles, Volume 1: Archostemata, Myxophaga, Adephaga, Polyphaga: Staphyliniformia., R.H.A. Jr and M.C. Thomas, eds. (CRC Press), pp. 272–418.
30. Thayer, M.K. (2005). Staphylinidae Latreille, 1802. In Handbuch Der Zoologie/Handbook of Zoology, Vol. IV (Arthropoda: Insecta), Part 38 Coleoptera, Beetles. Volume 1: Morphology and Systematics (Archostemata, Adephaga, Myxophaga, Polyphaga Partim), R.G. Beutel and R.A.B. Leschen, eds. (Walter de Gruyter), pp. 296–344.
31. Ashe, J.S. (2005). Phylogeny of the tachyporine group subfamilies and 'basal' lineages of the Aleocharinae (Coleoptera: Staphylinidae) based on larval and adult characteristics. Syst. Entomol. *30*, 3–37. https://doi.org/10.1111/j.1365-3113.2004.00258.x.
32. Yamamoto, S., and Maruyama, M. (2017). Phylogeny of the rove beetle tribe Gymnusini sensu n. (Coleoptera: Staphylinidae: Aleocharinae): implications for the early branching events of the subfamily. Syst. Entomol. *43*, 183–199. https://doi.org/10.1111/syen.12267.
33. 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.
34. Brückner, A., Badroos, J.M., Learsch, R.W., Yousefelahiyeh, M., Kitchen, S.A., and Parker, J. (2021). Evolutionary assembly of cooperating cell types in an animal chemical defense system. Cell *184*, 6138–6156.e28. https://doi.org/10.1016/j.cell.2021.11.014.
35. Jordan, K. (1913). Zur Morphologie und Biologie der myrmecophilen Gattungen *Lomechusa* und *Atemeles* und einiger verwandter Formen. Zeitschrift für wissenschaftliche Zoologie *107*, 346–386.
36. Donisthorpe, H. (1927). The Guests of British Ants: Their Habits and Life Histories (George Routledge & Sons).
37. Brand, J.M., Blum, M.S., Fales, H.M., and Pasteels, J.M. (1973). The chemistry of the defensive secretion of the beetle, *Drusilla canaliculata*.

J. Insect Physiol. *19*, 369–382. https://doi.org/10.1016/0022-1910(73)90112-1.

38. Hölldobler, B., Möglich, M., and Maschwitz, U. (1981). Myrmecophilic Relationship of *Pella* (Coleoptera: Staphylinidae) to *Lasius fuliginosus* (Hymenoptera: Formicidae). Psyche: J. Entomol. *88*, 347–374. https://doi.org/10.1155/1981/75317.
39. Stoeffler, M., Tolasch, T., and Steidle, J.L.M. (2011). Three beetles—three concepts. Different defensive strategies of congeneric myrmecophilous beetles. Behav. Ecol. Sociobiol. *65*, 1605–1613. https://doi.org/10.1007/s00265-011-1171-9.
40. Parker, J. (2016). Myrmecophily in beetles (Coleoptera): evolutionary patterns and biological mechanisms. Myrmecological News *22*, 65–108.
41. Parker, J., and Kronauer, D.J.C. (2021). How ants shape biodiversity. Curr. Biol. *31*, R1208–R1214. https://doi.org/10.1016/j.cub.2021.08.015.
42. Hölldobler, B. (1970). Zur Physiologie der Gast-Wirt-Beziehungen (Myrmecophilie) bei Ameisen. II. Das Gastverhältnis des imaginalen *Atemeles pubicollis* Bris. (Col. Staphylinidae) zu *Myrmica* und *Formica* (Hym. Formicidae). Zeitschrift für vergleichende Physiologie *66*, 215–250. https://doi.org/10.1007/bf00297780.
43. Kistner, D.H. (1979). Social and evolutionary significance of social insect symbionts. In Social Insects, *1*, H.R. Hermann, ed., pp. 339–413.
44. Maruyama, M., and Parker, J. (2017). Deep-Time Convergence in Rove Beetle Symbionts of Army Ants. Curr. Biol. *27*, 920–926. https://doi.org/10.1016/j.cub.2017.02.030.
45. Naragon, T.H., Wagner, J.M., and Parker, J. (2022). Parallel evolutionary paths of rove beetle myrmecophiles: replaying a deep-time tape of life. Curr. Opin. Insect Sci. *51*, 100903. https://doi.org/10.1016/j.cois.2022.100903.
46. Kistner, D.H., and Blum, M.S. (1971). Alarm Pheromone of *Lasius (Dendrolasius) Spathepus* (Hymenoptera: Formicidae) and its Possible Mimicry by Two Species of *Pella* (Coleoptera: Staphylinidae), *64* (Annals of the Entomological Society of America), pp. 589–594.
47. Stoeffler, M., Maier, T.S., Tolasch, T., and Steidle, J.L.M. (2007). Foreign-language skills in rove-beetles? Evidence for chemical mimicry of ant alarm pheromones in myrmecophilous *Pella* beetles (Coleoptera: Staphylinidae). J. Chem. Ecol. *33*, 1382–1392. https://doi.org/10.1007/s10886-007-9315-0.
48. Stoeffler, M., Boettinger, L., Tolasch, T., and Steidle, J.L.M. (2013). The Tergal Gland Secretion of the Two Rare Myrmecophilous Species *Zyras collaris* and *Z. haworthi* (Coleoptera: Staphylinidae) and the Effect on *Lasius fuliginosus*. Psyche: J. Entomol. *2013*, 1–5. https://doi.org/10.1155/2013/601073.
49. Miller, A.H., Stroud, J.T., and Losos, J.B. (2023). The ecology and evolution of key innovations. Trends Ecol. Evol. *38*, 122–131. https://doi.org/10.1016/j.tree.2022.09.005.
50. 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 *9*, 987–991. https://doi.org/10.1534/g3.119.300584.
51. Pflug, J.M., Holmes, V.R., Burrus, C., Johnston, J.S., and Maddison, D.R. (2020). Measuring Genome Sizes Using Read-Depth, k-mers, and Flow Cytometry: Methodological Comparisons in Beetles (Coleoptera). G3 (Bethesda). *10*, 3047–3060. https://doi.org/10.1534/g3.120.401028.
52. Mestrovic, N., Plohl, M., Mravinac, B., and Ugarkovic, D. (1998). Evolution of satellite DNAs from the genus *Palorus*–experimental evidence for the “library” hypothesis. Mol. Biol. Evol. *15*, 1062–1068. https://doi.org/10.1093/oxfordjournals.molbev.a026005.
53. Kasinathan, S., and Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. Mol. Biol. Evol. *35*, 949–962. https://doi.org/10.1093/molbev/msy010.
54. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744–757.e24. https://doi.org/10.1016/j.cell.2018.05.024.
55. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199.
56. Bracewell, R., Tran, A., Chatla, K., and Bachtrog, D. (2023). Sex chromosome evolution in beetles. Preprint at bioRxiv. https://doi.org/10.1101/2023.01.18.524646.
57. Osswald, J., Bachmann, L., and Gusarov, V.I. (2013). Molecular phylogeny of the beetle tribe Oxypodini (Coleoptera: Staphylinidae: Aleocharinae). Syst. Entomol. *38*, 507–522. https://doi.org/10.1111/syen.12011.
58. Orlov, I., Leschen, R.A.B., Żyła, D., and Solodovnikov, A. (2020). Total-evidence backbone phylogeny of Aleocharinae (Coleoptera: Staphylinidae). Cladistics *37*, 343–374. https://doi.org/10.1111/cla.12444.
59. Kistner, D.H., and Jacobson, H.R. (1990). Cladistic analysis and taxonomic revision of the ecitophilous tribe Ecitocharini with studies of their behaviour and evolution (Coleoptera, Staphylinidae, Aleocharinae). Sociobiology *17*, 333–480.
60. Kistner, D.H., Jensen, E.A., and Jacobson, H.R. (2002). A new genus and two new species of myrmecophilous Staphylinidae found with *Liometopum* in California (Coleoptera; Hymenoptera: Formicidae). Sociobiology *39*, 291–305.
61. Orlov, I., Newton, A.F., and Solodovnikov, A. (2021). Phylogenetic review of the tribal system of Aleocharinae, a mega-lineage of terrestrial arthropods in need of reclassification. J. Zool. Syst. Evol. Res. *59*, 1903–1938. https://doi.org/10.1111/jzs.12524.
62. Ahn, K.J., and Ashe, J.S. (2004). Phylogeny of the Myllaenini and related taxa (Coleoptera: Staphylinidae: Aleocharinae). Cladistics *20*, 123–138. https://doi.org/10.1111/j.1096-0031.2004.00012.x.
63. Elven, H., Bachmann, L., and Gusarov, V.I. (2010). Phylogeny of the tribe Athetini (Coleoptera: Staphylinidae) inferred from mitochondrial and nuclear sequence data. Mol. Phylogenet. Evol. *57*, 84–100. https://doi.org/10.1016/j.ympev.2010.05.023.
64. Elven, H., Bachmann, L., and Gusarov, V.I. (2012). Molecular phylogeny of the Athetini-Lomechusini-Ecitocharini clade of aleocharine rove beetles (Insecta). Zool. Scr. *41*, 617–636. https://doi.org/10.1111/j.1463-6409.2012.00553.x.
65. Grimaldi, D., and Agosti, D. (2000). A formicine in New Jersey Cretaceous amber (Hymenoptera: Formicidae) and early evolution of the ants. Proc. Natl. Acad. Sci. USA *97*, 13678–13683. https://doi.org/10.1073/pnas.240452097.
66. Engel, M.S., Grimaldi, D.A., and Krishna, K. (2009). Termites (Isoptera): Their Phylogeny, Classification, and Rise to Ecological Dominance. Am. Mus. Novit. *3650*, 1–27. https://doi.org/10.1206/651.1.
67. Barden, P. (2016). Fossil ants (Hymenoptera: Formicidae): ancient diversity and the rise of modern lineages. Myrmecological News *24*, 1–30.
68. Ibarra, Y., and Blair, N.T. (2013). Benzoquinone Reveals a Cysteine-Dependent Desensitization Mechanism of TRPA1. Mol. Pharmacol. *83*, 1120–1132. https://doi.org/10.1124/mol.112.084194.
69. Dettner, K. (1991). Solvent-dependent Variability of Effectiveness of Quinone-defensive Systems of Oxytelinae Beetles (Coleoptera: Staphylinidae). Entomol. Gen. *15*, 275–292. https://doi.org/10.1127/entom.gen/15/1991/275.
70. 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. https://doi.org/10.1016/0020-1790(84)90093-3.
71. Peschke, K., and Metzler, M. (1982). Defensive and pheromonal secretion of the tergal gland of *Aleochara curtula*: I. The chemical composition. J. Chem. Ecol. *8*, 773–783. https://doi.org/10.1007/bf00988318.

72. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., and Storey, J.D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics *28*, 882–883. https://doi.org/10.1093/bioinformatics/bts034.
73. Watkins, P.A. (1997). Fatty acid activation. Prog. Lipid Res. *36*, 55–83. https://doi.org/10.1016/s0163-7827(97)00004-0.
74. 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. https://doi.org/10.1093/gbe/evz094.
75. Lutsenko, S., and Petris, M.J. (2003). Function and Regulation of the Mammalian Copper-transporting ATPases: Insights from Biochemical and Cell Biological Approaches. J. Membr. Biol. *191*, 1–12. https://doi.org/10.1007/s00232-002-1040-6.
76. Norgate, M., Lee, E., Southon, A., Farlow, A., Batterham, P., Camakaris, J., and Burke, R. (2006). Essential Roles in Development and Pigmentation for the *Drosophila* Copper Transporter DmATP7. Mol. Biol. Cell *17*, 475–484. https://doi.org/10.1091/mbc.e05-06-0492.
77. 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. Preprint at bioRxiv. https://doi.org/10.1101/2020.12.11.421891.
78. Brückner, A., and Parker, J. (2020). Molecular evolution of gland cell types and chemical interactions in animals. J. Exp. Biol. *223*, jeb211938. https://doi.org/10.1242/jeb.211938.
79. Bowles, D., Lim, E.-K., Poppenberger, B., and Vaistij, F.E. (2006). Glycosyltransferases of lipophilic small molecules. Plant Biol. *57*, 567–597. https://doi.org/10.1146/annurev.arplant.57.032905.105429.
80. Beran, F., Köllner, T.G., Gershenzon, J., and Tholl, D. (2019). Chemical convergence between plants and insects: biosynthetic origins and functions of common secondary metabolites. New Phytol. *242*, 4157. https://doi.org/10.1111/nph.15718.
81. Morant, A.V., Jørgensen, K., Jørgensen, C., Paquette, S.M., Sánchez-Pérez, R., Møller, B.L., and Bak, S. (2008). β-Glucosidases as detonators of plant chemical defense. Phytochemistry *69*, 1795–1813. https://doi.org/10.1016/j.phytochem.2008.03.006.
82. Happ, G.M. (1968). Quinone and hydrocarbon production in the defensive glands of *Eleodes longicollis* and *Tribolium castaneum* (Coleoptera, Tenebrionidae). J. Insect Physiol. *14*, 1821–1837. https://doi.org/10.1016/0022-1910(68)90214-x.
83. Kaplowitz, N. (1980). Physiological significance of glutathione S-transferases. Am. J. Physiol. *239*, G439–G444. https://doi.org/10.1152/ajpgi.1980.239.6.g439.
84. Rowland, A., Miners, J.O., and Mackenzie, P.I. (2013). The UDP-glucuronosyltransferases: Their role in drug metabolism and detoxification. Int. J. Biochem. Cell Biol. *45*, 1121–1132. https://doi.org/10.1016/j.biocel.2013.02.019.
85. Kawamukai, M. (2016). Biosynthesis of coenzyme Q in eukaryotes. Biosci. Biotechnol. Biochem. *80*, 23–33. https://doi.org/10.1080/09168451.2015.1065172.
86. 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. https://doi.org/10.1073/pnas.0504982102.
87. 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. https://doi.org/10.1016/j.ibmb.2019.03.007.
88. Jeon, J.-R., Baldrian, P., Murugesan, K., and Chang, Y.-S. (2012). Laccase-catalysed oxidations of naturally occurring phenols: from *in vivo* biosynthetic pathways to green synthetic applications. Microb. Biotechnol. *5*, 318–332. https://doi.org/10.1111/j.1751-7915.2011.00273.x.
89. Shimoda, T., Shinkaji, N., and Amano, H. (1997). Prey stage preference and feeding behaviour of *Oligota kashmirica benefica* (Coleoptera: Staphylinidae), an insect predator of the spider mite *Tetranychus urticae* (Acari: Tetranychidae). Exp. Appl. Acarol. *21*, 665–675. https://doi.org/10.1007/bf02803509.
90. Perumalsamy, K., Selvasundaram, R., Roobakkumar, A., Rahman, V.J., Babu, A., and Muraleedharan, N.N. (2009). Life table and predation of *Oligota pygmaea* (Coleoptera: Staphylinidae) a major predator of the red spider mite, *Oligonychus coffeae* (Acarina: Tetranychidae) infesting tea. Biol. Control *51*, 96–101. https://doi.org/10.1016/j.biocontrol.2009.07.002.
91. Makki, R., Cinnamon, E., and Gould, A.P. (2014). The Development and Functions of Oenocytes. Annu. Rev. Entomol. *59*, 405–425. https://doi.org/10.1146/annurev-ento-011613-162056.
92. Blomquist, G.J., and Ginzel, M.D. (2021). Chemical Ecology, Biochemistry, and Molecular Biology of Insect Hydrocarbons. Annu. Rev. Entomol. *66*, 45–60. https://doi.org/10.1146/annurev-ento-031620-071754.
93. Mori, N., Kuwahara, Y., and Kurosa, K. (1998). Rosefuran: The Sex Pheromone of an Acarid Mite, *Caloglyphus* sp. J. Chem. Ecol. *24*, 1771–1779. https://doi.org/10.1023/a:1022399331397.
94. Raspotnig, G. (2006). Chemical alarm and defence in the oribatid mite *Collohmannia gigantea* (Acari: Oribatida). Exp. Appl. Acarol. *39*, 177–194. https://doi.org/10.1007/s10493-006-9015-4.
95. Nishimura, K., Shimizu, N., Mori, N., and Kuwahara, Y. (2002). Chemical ecology of astigmatid mites. LXIV. The alarm pheromone neral functions as an attractant *in Schwiebea elongata* (Banks) (Acari: Acaridae). Appl. Entomol. Zool. *37*, 13–18. https://doi.org/10.1303/aez.2002.13.
96. Curtis, R.F., Hobson-Frohock, A., Fenwick, G.R., and Berreen, J.M. (1981). Volatile compounds from the mite *Acarus siro* L. in food. J. Stored Prod. Res. *17*, 197–203. https://doi.org/10.1016/0022-474x(81)90006-0.
97. LEAL, W.S., NAKANO, Y., KUWAHARA, Y., NAKAO, H., and SUZUKI, T. (1988). Pheromone Study of Acarid Mites XVII, Identification of 2-Hydroxy-6-Methyl Benzaldehyde as the Alarm Pheromone of the Acarid Mite *Tyrophagus perniciosus* (Acarina:Acaridae), and its Distribution among Related Mites. Appl. Entomol. Zool. *23*, 422–427. https://doi.org/10.1303/aez.23.422.
98. SAKATA, T., and KUWAHARA, Y. (2001). Structural Elucidation and Synthesis of 3-Hydroxybenzene-1,2-dicarbaldehyde from Astigmatid Mites. Biosci. Biotechnol. Biochem. *65*, 2315–2317. https://doi.org/10.1271/bbb.65.2315.
99. Gould, S.J., and Lewontin, R.C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc. R. Soc. Lond. B Biol. Sci. *205*, 581–598. https://doi.org/10.1098/rspb.1979.0086.
100. Barrett, R.D.H., and Hoekstra, H.E. (2011). Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. *12*, 767–780. https://doi.org/10.1038/nrg3015.
101. Hölldobler, B., and Kwapich, C.L. (2019). Behavior and exocrine glands in the myrmecophilous beetle *Dinarda dentata* (Gravenhorst, 1806) (Coleoptera: Staphylinidae: Aleocharinae). PLoS One *14*, e0210524-22. https://doi.org/10.1371/journal.pone.0210524.
102. Hölldobler, B., Kwapich, C.L., and Haight, K.L. (2018). Behavior and exocrine glands in the myrmecophilous beetle *Lomechusoides strumosus* (Fabricius, 1775) (formerly called *Lomechusa strumosa*) (Coleoptera: Staphylinidae: Aleocharinae). PLoS One *13*, e0200309-38. https://doi.org/10.1371/journal.pone.0200309.
103. Hölldobler, B., and Kwapich, C. (2022). The Guests of Ants (Harvard University Press).
104. Hölldobler, B. (1967). Zur Physiologie der Gast-Wirt-Beziehungen (Myrmecophilie) bei Ameisen. I. Das Gastverhältnis der *Atemeles*- und *Lomechusa*-Larven (Col. Staphylinidae) zu *Formica* (Hym. Formicidae). Zeitschrift für vergleichende Physiologie *56*, 1–21. https://doi.org/10.1007/bf00333561.

105. Maruyama, M., Akino, T., Hashim, R., and Komatsu, T. (2009). Behavior and cuticular hydrocarbons of myrmecophilous insects (Coleoptera: Staphylinidae; Diptera: Phoridae; Thysanura) associated with Asian *Aenictus* army ants (Hymenoptera; Formicidae). Sociobiology *54*, 19–35.

106. Akino, T. (2002). Chemical camouflage by myrmecophilous beetles *Zyras comes* (Coleoptera: Staphylinidae) and *Diartiger fossulatus* (Coleoptera: Pselaphidae) to be integrated into the nest of *Lasius fuliginosus* (Hymenoptera: Formicidae). Chemoecology *12*, 83–89.

107. Beran, F., Rahfeld, P., Luck, K., Nagel, R., Vogel, H., Wielsch, N., Irmisch, S., Ramasamy, S., Gershenzon, J., Heckel, D.G., and Köllner, T.G. (2016). Novel family of terpene synthases evolved from trans-isoprenyl diphosphate synthases in a flea beetle. Proc. Natl. Acad. Sci. USA *113*, 2922–2927. https://doi.org/10.1073/pnas.1523468113.

108. Darragh, K., Orteu, A., Black, D., Byers, K.J.R.P., Szczerbowski, D., Warren, I.A., Rastas, P., Pinharanda, A., Davey, J.W., Fernanda Garza, S., et al. (2021). A novel terpene synthase controls differences in anti-aphrodisiac pheromone production between closely related *Heliconius* butterflies. PLoS Biol. *19*, e3001022. https://doi.org/10.1371/journal.pbio.3001022.

109. Rebholz, Z., Lancaster, J., Larose, H., Khrimian, A., Luck, K., Sparks, M.E., Gendreau, K.L., Shewade, L., Köllner, T.G., Weber, D.C., et al. (2023). Ancient origin and conserved gene function in terpene pheromone and defense evolution of stink bugs and hemipteran insects. Insect Biochem Molec *152*, 103879. https://doi.org/10.1016/j.ibmb.2022.103879.

110. Toprak, U., Hegedus, D., Doğan, C., and Güney, G. (2020). A journey into the world of insect lipid metabolism. Arch Insect Biochem *104*, e21682. https://doi.org/10.1002/arch.21682.

111. Arrese, E.L., and Soulages, J.L. (2010). Insect Fat Body: Energy, Metabolism, and Regulation. Annu. Rev. Entomol. *55*, 207–225. https://doi.org/10.1146/annurev-ento-112408-085356.

112. Jacobson, H.R., and Kistner, D.H. (1991). Cladistic study, taxonomic restructuring, and revision of the myrmecophilous tribe Leptanillophilini with comments on its evolution and host relationships (Coleoptera: Staphylinidae; Hymenoptera: Formicidae). Sociobiology *18*, 1–150.

113. Kistner, D.H. (1993). Cladistic analysis, taxonomic restructuring and revision of the Old World genera formerly classified as Dorylomimini with comments on their evolution and behavior (Coleoptera: Staphylinidae). Sociobiology *22*, 147–383.

114. Beeren, C. von, Brueckner, A., Maruyama, M., Burke, G., Wieschollek, J., and Kronauer, D.J.C. (2018). Chemical and behavioral integration of army ant-associated rove beetles - a comparison between specialists and generalists. Front. Zool. *15*, 1–15. https://doi.org/10.1186/s12983-018-0249-x.

115. Parker, J. (2024). Organ Evolution: Emergence of Multicellular Function. Annu. Rev. Cell Dev. Biol. https://doi.org/10.1146/annurev-cellbio-111822-121620.

116. Des Marais, D.L., and Rausher, M.D. (2008). Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature *454*, 762–765. https://doi.org/10.1038/nature07092.

117. Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. *11*, 97–108. https://doi.org/10.1038/nrg2689.

118. Hobert, O., and Kratsios, P. (2019). Neuronal identity control by terminal selectors in worms, flies, and chordates. Curr. Opin. Neurobiol. *56*, 97–105. https://doi.org/10.1016/j.conb.2018.12.006.

119. 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. Preprint at bioRxiv. https://doi.org/10.1101/198945.

120. Dettner, K. (1993). 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. https://doi.org/10.1007/bf00984880.

121. Oakley, T.H., and Speiser, D.I. (2015). How Complexity Originates: The Evolution of Animal Eyes. Annu. Rev. Ecol. Evol. Syst. *46*, 237–260. https://doi.org/10.1146/annurev-ecolsys-110512-135907.

122. Babonis, L.S., Enjolras, C., Ryan, J.F., and Martindale, M.Q. (2022). A novel regulatory gene promotes novel cell fate by suppressing ancestral fate in the sea anemone *Nematostella vectensis*. Proc. Natl. Acad. Sci. USA *119*, e2113701119. https://doi.org/10.1073/pnas.2113701119.

123. Jeffries, M.J., and Lawton, J.H. (1984). Enemy free space and the structure of ecological communities. Biol. J. Linn. Soc. *23*, 269–286. https://doi.org/10.1111/j.1095-8312.1984.tb00145.x.

124. Quinodoz, S.A., Bhat, P., Chovanec, P., Jachowicz, J.W., Ollikainen, N., Detmar, E., Soehalim, E., and Guttman, M. (2022). SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. Nat. Protoc. *17*, 36–75. https://doi.org/10.1038/s41596-021-00633-y.

125. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data.

126. Chikhi, R., and Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. Bioinformatics *30*, 31–37. https://doi.org/10.1093/bioinformatics/btt310.

127. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

128. Hozza, M., Vinař, T., and Brejová, B. (2015). How Big is that Genome? Estimating Genome Size and Coverage from k-mer Abundance Spectra. Lect Notes Comput Sc, 199–209. https://doi.org/10.1007/978-3-319-23826-5_20.

129. Sun, H., Ding, J., Piednoël, M., and Schneeberger, K. (2018). findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. Bioinformatics *34*, 550–557. https://doi.org/10.1093/bioinformatics/btx637.

130. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

131. Ranallo-Benavidez, T.R., Jaron, K.S., and Schatz, M.C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. *11*, 1432. https://doi.org/10.1038/s41467-020-14998-3.

132. 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. https://doi.org/10.1016/j.ymeth.2016.02.020.

133. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. F1000Res. *6*, 1287. https://doi.org/10.12688/f1000research.12232.1.

134. Pryszcz, L.P., and Gabaldón, T. (2016). Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic Acids Res. *44*, e113. https://doi.org/10.1093/nar/gkw294.

135. 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. https://doi.org/10.1186/2047-217x-1-18.

136. Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinf. *15*, 211. https://doi.org/10.1186/1471-2105-15-211.

137. Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Methods *17*, 155–158. https://doi.org/10.1038/s41592-019-0669-3.

138. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via

adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736. https://doi.org/10.1101/gr.215087.116.

139. Chakraborty, M., Baldwin-Brown, J.G., Long, A.D., and Emerson, J.J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Res. *44*, e147. https://doi.org/10.1093/nar/gkw654.
140. Xu, G.-C., Xu, T.-J., Zhu, R., Zhang, Y., Li, S.-Q., Wang, H.-W., and Li, J.-T. (2019). LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. GigaScience *8*, giy157. https://doi.org/10.1093/gigascience/giy157.
141. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLoS One *9*, e112963. https://doi.org/10.1371/journal.pone.0112963.
142. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746. https://doi.org/10.1101/gr.214270.116.
143. Roach, M.J., Schmidt, S.A., and Borneman, A.R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinf. *19*, 460. https://doi.org/10.1186/s12859-018-2485-7.
144. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
145. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
146. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.
147. Renschler, G., Richard, G., Valsecchi, C.I.K., Toscano, S., Arrigoni, L., Ramírez, F., and Akhtar, A. (2019). Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. Genes Dev. *33*, 1591–1612. https://doi.org/10.1101/gad.328971.119.
148. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
149. Smit, A.F., and Hubley, R. (2008). RepeatModeler Open-1.0.
150. 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 Bioinf. *19*, 348. https://doi.org/10.1186/s12859-018-2376-y.
151. Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584. https://doi.org/10.7717/peerj.2584.
152. Goubert, C., Modolo, L., Vieira, C., ValienteMoro, C., Mavingui, P., and Boulesteix, M. (2015). De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti). Genome Biol. Evol. *7*, 1192–1205. https://doi.org/10.1093/gbe/evv050.
153. Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat. Protoc. *15*, 3745–3776. https://doi.org/10.1038/s41596-020-0400-y.
154. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
155. Negm, S., Greenberg, A., Larracuente, A.M., and Sproul, J.S. (2021). RepeatProfiler: A pipeline for visualization and comparative analysis of repetitive DNA profiles. Mol. Ecol. Resour. *21*, 969–981. https://doi.org/10.1111/1755-0998.13305.
156. Gao, Y., Liu, B., Wang, Y., and Xing, Y. (2019). TideHunter: efficient and sensitive tandem repeat detection from noisy long-reads using seed-and-chain. Bioinformatics *35*, i200–i207. https://doi.org/10.1093/bioinformatics/btz376.
157. Seibt, K.M., Schmidt, T., and Heitkam, T. (2018). FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. Bioinformatics *34*, 3575–3577. https://doi.org/10.1093/bioinformatics/bty395.
158. Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. *33*, 6494–6506. https://doi.org/10.1093/nar/gki937.
159. 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. https://doi.org/10.1093/nargab/lqaa108.
160. 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. https://doi.org/10.1186/gb-2008-9-1-r7.
161. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinf. *6*, 31. https://doi.org/10.1186/1471-2105-6-31.
162. Keilwagen, J., Hartung, F., and Grau, J. (2019). Gene Prediction, Methods and Protocols. Methods Mol. Biol. *1962*, 161–177. https://doi.org/10.1007/978-1-4939-9173-0_9.
163. 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. https://doi.org/10.1093/bioinformatics/bts635.
164. 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. https://doi.org/10.1038/nprot.2013.084.
165. Wu, T.D., and Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics *21*, 1859–1875. https://doi.org/10.1093/bioinformatics/bti310.
166. Kent, W.J. (2002). BLAT—The BLAST-Like Alignment Tool. Genome Res. *12*, 656–664. https://doi.org/10.1101/gr.229202.
167. Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics *37*, 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016.
168. Taylor, W.R., and Orengo, C.A. (1989). Protein structure alignment. J. Mol. Biol. *208*, 1–22. https://doi.org/10.1016/0022-2836(89)90084-3.
169. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. *34*, 2115–2122. https://doi.org/10.1093/molbev/msx148.
170. 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. https://doi.org/10.1093/bioinformatics/bts565.
171. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.
172. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

173. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS One *5*, e9490. https://doi.org/10.1371/journal.pone.0009490.

174. Misof, B., Meyer, B., von Reumont, B.M., Kück, P., Misof, K., and Meusemann, K. (2013). Selecting informative subsets of sparse supermatrices increases the chance to find correct trees. BMC Bioinf. *14*, 348. https://doi.org/10.1186/1471-2105-14-348.

175. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods *14*, 587–589. https://doi.org/10.1038/nmeth.4285.

176. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

177. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol. Biol. Evol. *35*, 518–522. https://doi.org/10.1093/molbev/msx281.

178. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinf. *19*, 153. https://doi.org/10.1186/s12859-018-2129-y.

179. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

180. Puttick, M.N. (2019). MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics *35*, 5321–5322. https://doi.org/10.1093/bioinformatics/btz554.

181. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. *34*, W435–W439. https://doi.org/10.1093/nar/gkl200.

182. Emms, D.M., and Kelly, S. (2022). SHOOT: phylogenetic gene search and ortholog inference. Genome Biol. *23*, 85. https://doi.org/10.1186/s13059-022-02652-8.

183. 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. https://doi.org/10.1093/molbev/mst010.

184. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2019). HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. *37*, 295–299. https://doi.org/10.1093/molbev/msz197.

185. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. *33*, 1635–1638. https://doi.org/10.1093/molbev/msw046.

186. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. *16*, 276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

187. Kirilenko, B.M., Munegowda, C., Osipova, E., Jebb, D., Sharma, V., Blumer, M., Morales, A., Ahmed, A.-W., Kontopoulos, D.-G., Hilgers, L., et al. (2022). TOGA integrates gene annotation with orthology inference at scale. Preprint at bioRxiv. https://doi.org/10.1101/2022.09.08.507143.

188. Harris, R.S. (2007). Improved Pairwise Alignment of Genomic DNA. PhD thesis (Pennsylvania State University).

189. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly *6*, 80–92. https://doi.org/10.4161/fly.19695.

190. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

191. Auwera, G.A.V. der, and O'Connor, B.D. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media)

192. Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12. https://doi.org/10.1186/gb-2004-5-2-r12.

193. Haas, B.J., Delcher, A.L., Wortman, J.R., and Salzberg, S.L. (2004). DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics *20*, 3643–3646. https://doi.org/10.1093/bioinformatics/bth397.

194. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics *20*, 289–290.

195. Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R. (1996). Computer Visualization of Three-Dimensional Image Data Using IMOD. J. Struct. Biol. *116*, 71–76. https://doi.org/10.1006/jsbi.1996.0013.

196. MASTRONARDE, D.N. (2008). Correction for non-perpendicularity of beam and tilt axis in tomographic reconstructions with the IMOD package. J. Microsc *230*, 212–217. https://doi.org/10.1111/j.1365-2818.2008.01977.x.

197. Mastronarde, D.N., and Held, S.R. (2017). Automated tilt series alignment and tomographic reconstruction in IMOD. J. Struct. Biol. *197*, 102–113. https://doi.org/10.1016/j.jsb.2016.07.011.

198. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

199. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

200. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. Omics J Integr Biology *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

201. Kolde, R., and Kolde, M.R. (2015). Package 'pheatmap.'. R package 1, p. 790.

202. Ebbert, D. (2019). A post hoc analysis for Pearson's chi-squared test for count data. R package version 0.1, p. 2.

203. Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. BMC Bioinform *17*, 483. https://doi.org/10.1186/s12859-016-1323-z.

204. Zhang, Y., Parmigiani, G., and Johnson, W.E. (2020). ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genom. Bioinform. *2*, lqaa078. https://doi.org/10.1093/nargab/lqaa078.

205. Carlson, M., and Pages, H. (2019). AnnotationForge: tools for building SQLite-based annotation data packages. R package version 1.

206. Team, R.C. (2022). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

207. Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. *34*, W609–W612. https://doi.org/10.1093/nar/gkl315.

208. Steenwyk, J.L., Goltz, D.C., Buida, T.J., Li, Y., Shen, X.-X., and Rokas, A. (2022). OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. PLoS Biol. *20*, e3001827. https://doi.org/10.1371/journal.pbio.3001827.

209. Gilbert, M.T.P., Moore, W., Melchior, L., and Worobey, M. (2007). DNA Extraction from Dry Museum Beetles without Conferring External Morphological Damage. PLoS One *2*, e272. https://doi.org/10.1371/journal.pone.0000272.

210. Johnston, J.S., Bernardini, A., and Hjelmen, C.E. (2018). Genome Size Estimation and Quantitative Cytogenetics in Insects. In Insect Genomics: Methods and Protocols Methods in Molecular Biology, pp. 15–26. https://doi.org/10.1007/978-1-4939-8775-7_2.

211. Karlsson, E., Lärkeryd, A., Sjödin, A., Forsman, M., and Stenberg, P. (2015). Scaffolding of a bacterial genome using MinION nanopore sequencing. Sci. Rep. *5*, 11996. https://doi.org/10.1038/srep11996.

212. Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun *9*, 189. https://doi.org/10.1038/s41467-017-02525-w.

213. Brückner, A., Barnett, A.A., Bhat, P., Antoshechkin, I.A., and Kitchen, S.A. (2022). Molecular evolutionary trends and biosynthesis pathways in the Oribatida revealed by the genome of Archegozetes longisetosus. Acarologia *62*, 532–573.

214. 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. https://doi.org/10.1186/s12862-018-1324-9.

215. Smit, A., Hubley, R., and Green, P. RepeatMasker Open-4.0

216. 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. https://doi.org/10.1038/nbt.1883.

217. Chernomor, O., von Haeseler, A., and Minh, B.Q. (2016). Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. Syst. Biol. *65*, 997–1008. https://doi.org/10.1093/sysbio/syw037.

218. Zhang, S.-Q., Che, L.-H., Li, Y., Liang, D., Pang, H., Ślipiński, A., and Zhang, P. (2018). Evolutionary history of Coleoptera revealed by extensive sampling of genes and species. Nat. Commun. *9*, 205. https://doi.org/10.1038/s41467-017-02644-4.

219. Cai, C., Tihelka, E., Giacomelli, M., Lawrence, J.F., Ślipiński, A., Kundrata, R., Yamamoto, S., Thayer, M.K., Newton, A.F., Leschen, R.A.B., et al. (2022). Integrated phylogenomics and fossil data illuminate the evolution of beetles. R. Soc. Open Sci. *9*, 211771. https://doi.org/10.1098/rsos.211771.

220. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. *67*, 901–904. https://doi.org/10.1093/sysbio/syy032.

221. Tupec, M., Buček, A., Janoušek, V., Vogel, H., Prchalova, D., Kindl, J., Pavlickova, T., Wenzelova, P., Jahn, U., Valterová, I., et al. (2019). Expansion of the fatty acyl reductase gene family shaped pheromone communication in Hymenoptera. Elife *8*, e39231. https://doi.org/10.7554/eLife.39231.

222. Oakeshott, J.G., Johnson, R.M., Berenbaum, M.R., Ranson, H., Cristino, A.S., and Claudianos, C. (2010). Metabolic enzymes associated with xenobiotic and chemosensory responses in Nasonia vitripennis. Insect Mol. Biol. *19*, 147–163. https://doi.org/10.1111/j.1365-2583.2009.00961.x.

223. Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. Mol. Biol. Evol. *32*, 1342–1353. https://doi.org/10.1093/molbev/msv022.

224. Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. Mol. Biol. Evol. *32*, 820–832. https://doi.org/10.1093/molbev/msu400.

225. Yang, Z., Kumar, S., and Nei, M. (1995). A new method of inference of ancestral nucleotide and amino acid sequences. Genetics *141*, 1641–1650. https://doi.org/10.1093/genetics/141.4.1641.

226. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223. https://doi.org/10.1111/j.2041-210x.2011.00169.x.

227. Rożek, M. (1994). A new chromosome preparation technique for Coleoptera (Insecta). Chromosome Res. *2*, 76–78. https://doi.org/10.1007/bf01539458.

228. Mastronarde, D.N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. *152*, 36–51. https://doi.org/10.1016/j.jsb.2005.07.007.

229. Kitchen, S.A., Crowder, C.M., Poole, A.Z., Weis, V.M., and Meyer, E. (2015). De Novo Assembly and Characterization of Four Anthozoan (Phylum Cnidaria) Transcriptomes. G3 (Bethesda). *5*, 2441–2452. https://doi.org/10.1534/g3.115.020164.

230. 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. https://doi.org/10.1242/dev.165753.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
BL21 competent *E. coli*	provided by C. VanDrisse, Caltech	N/A
Biological samples		
Adinopsis sp.	wild caught by K. Taro Eldredge in Rhode Island, USA	N/A
Aleochara nigra	Peschke Laboratory (University of Freiburg)	N/A
Aleochara sp.	wild caught by J. Parker in California, USA	N/A
Atheta pasadenae	wild caught by J. Parker in California, USA	N/A
Coproporus ventriculus	wild caught by J. Parker in New York, USA	N/A
Cypha longicornis	wild caught by C. Barnes in United Kingdom	N/A
Dalotia coriaria	Applied Bionomics (Canada)	https://www.appliedbio-nomics.com/products/dalotia/
Deinopsis erosa	wild caught by K. Taro Eldredge in Netherlands	N/A
Drosophila melanogaster	Dickinson Laboratory (Caltech)	N/A
Drusilla canaliculata	wild caught by J. Parker	N/A
Earota dentata	wild caught by K. Taro Eldredge	N/A
Ecitodaemon sp.	wild caught by M. Maruyama in Peru	N/A
Ecitomorpha nevermanni	wild caught by C. von Beeren in Costa Rica	N/A
Ecitophya simulans	wild caught by C. von Beeren in Costa Rica	N/A
Falagria sp.	wild caught by K. Taro Eldredge in Massachusetts, USA	N/A
Geostiba sp.	wild caught by M. Caterino in North Carolina, USA	N/A
Gymnusa sp.	wild caught by J. Parker in Canada	N/A
Holobus sp.	wild caught by T. H. Naragon California, USA	N/A
Leptusa sp.	wild caught by J. Parker in California, USA	N/A
Liometoxenus newtonarum	wild caught by J. Parker in California, USA	N/A
Lissagria laeviuscula	wild caught by J. M. Wagner California, USA	N/A
Myllaena sp.	wild caught by M. Caterino in South Carolina, USA	N/A
Oligota sp.	wild caught by T. H. Naragon in California, USA	N/A
Oxypoda opaca	wild caught by K. Taro Eldredge in Massachusetts, USA	N/A
Tachinus sp.	wild caught by J. Parker in California, USA	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sepedophilus sp.	wild caught by J. Parker in California, USA	N/A
Chemicals, peptides, and recombinant proteins		
13C6-tyrosine >99 atom % 13C, 99% (CP)	Sigma-Aldrich	Cat# 489794
RNaseA	Qiagen	Cat# 19101
ExoSAP-IT	Thermo Fisher Scientific	Cat# 78200.200.UL
DLE-1	Provided by HistoGenetics, NY, USA	N/A
disuccinimidyl glutarate	Thermo Fisher Scientific	Cat# 20593
Phosphate Buffered Saline	Sigma-Aldrich	Cat#P4417
Formaldehyde ampules, 16%, methanol-free	Thermo Scientific Pierce	Cat# PI28908
Glycine, >99%	Sigma-Aldrich	Cat# G7403
RNase-Free BSA	American Bio	Cat # AB01243
HEPES buffer, pH 7.4, 1 M	Teknova	Cat# H1030
EDTA, 0.5 M, pH 8.0	Thermo Fisher Scientific	Cat# 15575020
Sodium chloride (NaCl)	Thermo Fisher Scientific	Cat# AM9759
Triton X-100	Sigma-Aldrich	Cat# T8787
NP-40 Surfact-Amps detergent	Thermo Fisher Scientific	Cat# 28324
glycerol	Sigma-Aldrich	Cat# G5516
EGTA, 0.5 M, pH 8.0	Fisher Scientific	Cat# 50255957
Sodium deoxycholate (DOC)	Sigma-Aldrich	Cat# D6750
N-Lauroylsarcosine sodium salt solution, 20% solution	Sigma-Aldrich	Cat# L7414
Manganese chloride ($MnCl_2$) solution, 1 M	Sigma-Aldrich	Cat# M1787
Calcium chloride ($CaCl_2$) solution, 1 M	Sigma-Aldrich	Cat# 21115
Tris-HCl pH 7.5	Thermo Fisher Scientific	Cat# 15567027
Sodium dodecyl sulfate (SDS)	Thermo Fisher Scientific	Cat# AM9820
Buffer RLT	Qiagen	Cat# 79216
NEBNext quick ligation reaction buffer	New England Biolabs	cat. no. B6058S
1,2-propanediol	Sigma-Aldrich	Cat# 398039
Instant Sticky-end Ligation Master Mix	New England Biolabs	Cat# M0370
Lithium chloride solution, 8 M (LiCl)	Sigma-Aldrich	Cat# L7026
Hydrochloric acid	VWR	Cat# 470301-260
Proteinase K, Molecular Biology Grade (ProK), 800 U/mL	New England Biolabs	Cat# P8107S
Protease cocktail inhibitor tablets	Sigma-Aldrich	Cat# 04693159001
hexane ReagentPlus, ≥99%	Sigma-Aldrich	Cat#139386
Diethyl pyrocarbonate (DEPC)	Sigma-Aldrich	Cat# 40718
tridecane	Sigma-Aldrich	Cat# 91490
2,5-dimethoxybenzaldehyde	Sigma-Aldrich	Cat# D130605
ethyl linoleate	Sigma-Aldrich	Cat# L1751
KaryoMAX™ Colcemid™ Solution in PBS	Gibco	Cat# 15212012
Potassium chloride (KCl)	Sigma-Aldrich	Cat# P3911
Acetic Acid glacial, *ReagentPlus*®, ≥99%	Sigma-Aldrich	Cat# A6283
Hoechst 33342	Thermo Fisher Scientific	Cat# 62249
VECTASHIELD® Antifade Mounting Medium with DAPI	Vector Laboratories	Cat# H-1200-10
Aqueous Glutaraldehyde EM Grade, 10%	Electron Microscopy Sciences	Cat# 16110
Sucrose, Reagent, A.C.S.	Electron Microscopy Sciences	Cat# 21600
sodium cacodylate trihydrate	Electron Microscopy Sciences	Cat# 11653
Ficoll® 400	Sigma-Aldrich	Cat# F8016

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Osmium tetroxide solution, 4% in H_2O	Sigma-Aldrich	Cat# 75632
uranyl acetate	Electron Microscopy Sciences	Cat# 22400
Acetone, Reagent Grade	Electron Microscopy Sciences	Cat# 10014
Epon-Araldite resin	Electron Microscopy Sciences	Cat# 14130
Lead(II) citrate tribasic trihydrate	Sigma-Aldrich	Cat# 15326
Trizol Reagent	Thermo Fisher Scintific	Cat# 15596026
Alexa 488- or Alexa 647-Wheat Germ Agglutinin Conjugate	Thermo Fisher	Cat# W11261 or Cat# W32466
ProLong Gold Antifade Mountant	Thermo Fisher Scientific	Cat# P36934
ABTS (2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid))	Sigma-Aldrich	Cat# 10102946001
1,4-hydroquinone ReagentPlus, 99%	Sigma-Aldrich	Cat# H17902
2-methyl-1,4-benzoquinone 98%	Sigma-Aldrich	Cat# 211311
2-methoxy-3-methy-1,4-hydroquinone	provided by A. Brückner (Brückner et al. 2021)[34]	N/A
2-Morpholinoethanesulfonic acid monohydrate EMPROVE EXPERT	Sigma-Aldrich	Cat# 1370740250
Copper (II) sulfate ReagentPlus, ≥99%	Sigma-Aldrich	Cat# C1297
Schneider's *Drosophila* medium	Thermo Fisher	Cat# 21720024
Isopropyl-β-D-thiogalactoside (IPTG)	Sigma-Aldrich	Cat# 10724815001
imidazole ReagentPlus, 99%	Sigma-Aldrich	Cat# I202
Halt™ Protease Inhibitor cocktail	Thermo Scientific	Cat# 78430
Ni-NTA resin	Sigma-Aldrich	Cat# 70666-4
SnakeSkin Dialysis Tubing, 10K MWCO, 16 mm	Thermo Fisher Scientific	Cat# 68100
urea, BioReagent, for molecular biology,	Sigma-Aldrich	Cat# U5378
DNase grade II, from bovine pancreas	Sigma-Aldrich	Cat# 10104159001
Critical commercial assays		
DNeasy Blood and Tissue extraction kit	Qiagen	Cat# 69504
Monarch PCR and DNA Cleanup kit	New England Biolabs	Cat# T1030S
Qubit 1X High Sensitivity dsDNA kit	Thermo Fisher Scientific	Cat# Q33230
Illumina TruSeq DNA	Illumina	Cat# FC-121-2001
NEBNext Ultra FS DNA library kit	New England Biolabs	Cat# E7805L
Bioanalyzer High Sensitivity DNA kit	Agilent Technologies	Cat# 5067-4626
MinION Nanopore vR9	Oxford Nanopore Technologies	Cat# FLO-MIN106D
MagAttract HMW DNA Kit	Qiagen	Cat# 67563
Turbo DNase	Thermo Fisher Scientific	Cat# AM2239
TOPO TA Cloning Kit	Thermo Fisher Scientific	Cat# 450641
MEGAclear Transcription Clean-Up kit	Thermo Fisher Scientific	Cat# AM1908
ZYMO *Quick*-RNA Tissue/Insect extraction kit	ZYMO Research	Cat# R2030
RNeasy Mini kit	Qiagen	Cat# 74104
Bioanalyzer High Sensitivity RNA Analysis kit	Agilent Technologies	Cat# 5067-1513
Illumina TruSeq RNA library kit	Illumina	Cat# RS-122-2001
NEBNext Single Cell/Low Input RNA Library Prep Kit	New England Biolabs	Cat# E6420L
MEGAscript T7 Transcription Kit	Thermo Fisher Scientific	Cat #AMB13345
Deposited data		
Dalotia coriaria Bionano Optipal Map data	This study	CaltechDATA: https://doi.org/10.22002/1914a-m9460
Dalotia genome assembly and gene predictions	This study	CaltechDATA: https://doi.org/10.22002/62xxb-mak64

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Other rove beetle genome assemblies and gene predictions	This study	CaltechDATA: https://doi.org/10.22002/k8sfv-dw648
Sequences and tree files	This study	CaltechDATA: https://doi.org/10.22002/cgsw0-9kk67
Selection test results	This study	CaltechDATA: https://doi.org/10.22002/gz6w6-g5355
Inactivating mutation associated files	This study	CaltechDATA: https://data.caltech.edu/records/6xjn1-e3085
Dalotia coriaria genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_025399875.2 SRR4301137
Ecitomorpha nevermanni genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_027574945.2 SRR5259840
Earota dentata genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_027574905.2 SRR5176873
Deinopsis erosa genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_027574845.2 SRR5176562
Coproporus ventriculus genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_027574865.2 SRR4301367
Ecitodaemon sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030557295.1 SRR23816754
Ecitophya simulans genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_027574965.2 SRR4301374
Oxypoda opaca genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264175.1 SRR23816753
Drusilla canaliculata genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_027574885.2 SRR5906249
Geostiba sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264215.1 SRR23816752
Myllaena sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264135.1 SRR23816751
Atheta pasadenae genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264155.1 SRR23816750
Leptusa sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264655.1 SRR23816749
Falagria sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030556245.1 SRR23816748
Lissagria laeviuscula genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264695.1 SRR23816747
Holobus sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030556065.1 SRR23816746
Aleochara nigra genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264675.1 SRR23816744
Adinopsis sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264715.1 SRR23816743
Gymnusa sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264735.1 SRR23816742
Cypha longicornis genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264615.1 SRR23816741
Aleochara sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264555.1 SRR23816854
Oligota sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264595.1 SRR23816775
Liometoxenus newtonarum genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264535.1 SRR23816853

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Tachinus sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264575.1 SRR15992418
Sepedophilus sp. genome assembly	This study; NCBI GenBank NCBI SRA database	GCA_030264515.1 SRR23816776
Aleochara bilineata genome assembly	NCBI GenBank, reassembly from this study at CaltechDATA	GCA_003054995.1; CaltechDATA: https://doi.org/10.22002/k8sfv-dw648
Anoplophora glabripennis genome assembly	NCBI RefSeq	GCF_000390285.2
Agrilus planipennis genome assembly	NCBI RefSeq	GCF_000699045.2
Aethina tumida genome assembly	NCBI RefSeq	GCF_001937115.1
Dendroctonus ponderosae genome assembly	NCBI RefSeq	GCF_000355655.1
Leptinotarsa decemlineata genome assembly	NCBI RefSeq	GCF_000500325.1
Nicrophorus vespilloides genome assembly	NCBI RefSeq	GCF_001412225.1
Onthophagus taurus genome assembly	NCBI RefSeq	GCF_000648695.1
Photinus pyralis genome assembly	NCBI RefSeq	GCF_008802855.1
Tribolium castaneum genome assembly	NCBI RefSeq	GCF_000002335.3
Philonthus cognatus genome assembly	NCBI GenBank	GCA_932526585.1
Ocypus olens genome assembly	NCBI GenBank	GCA_910593695.1
Aleochara sp. tissue-specific RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816793-SRR23816799, SRR23816801-SRR23816807
Aleochara sp. bulk whole organism RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816847
Dalotia coriaria bulk tissue and sex-specific RNAseq	This study, NCBI SRA Database, see Data S1L	SRR23816756-SRR23816758, SRR23816773, SRR23816774, SRR23816777, SRR23816778, SRR23816779, SRR23816780, SRR23816782- SRR23816791
Dalotia coriaria tissue-specific RNASeq	Brückner et al. 2021; NCBI SRA Database, see Data S1L	SRR13865081-SRR13865085, SRR13865092-SRR13865117
Holobus sp. bulk whole organism RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816851
Holobus sp. tissue-specific RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816763, SRR23816765, SRR23816767, SRR23816769, SRR23816771, SRR23816840, SRR23816841, SRR23816842, SRR23816843, SRR23816845, SRR23816846
Liometoxenus newtonarum tissue-specific RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816808-SRR23816810, SRR23816812-SRR23816821, SRR23816823-SRR23816828
Liometoxenus newtonarum bulk RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816848, SRR23816849
Oligota sp. tissue-specific RNASeq	This study, NCBI SRA Database, see Data S1L	SRR23816829-SRR23816832, SRR23816834-SRR23816839
Dalotia coriaria SPRITE raw data	This study, NCBI SRA Database, see Data S1L	SRR23816745, SRR23816755, SRR23816759, SRR23816760, SRR23816761, SRR23816762, SRR23816764, SRR23816766, SRR23816768, SRR23816770, SRR23816772, SRR23816781, SRR23816792, SRR23816800, SRR23816811, SRR23816822, SRR23816833, SRR23816844, SRR23816850, SRR23816855
Dalotia coriaria Nanopore raw data	This study, NCBI SRA Database, see Data S1L	SRR23816856
Dalotia coriaria WGS2	This study, NCBI SRA Database, see Data S1L	SRR23816852

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Indexing SPRITE Library Amplification primers	Quinodoz et al.[124]; Integrated DNA Technologies, Inc	N/A
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1)	New England Biolabs	Cat# E7600L
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 2)	New England Biolabs	Cat# E7780L
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 3)	New England Biolabs	Cat# E7710S
In Situ HCR probes from Molecular Technologies	Brückner et al. 2021 and This Study, see Data S1N	N/A
In Situ HCR probes from IDT	This Study; see Data S1N	N/A
Dalotia bubblegum (*bgm*) F	This Study; Integrated DNA Technologies, Inc	5′-TAATACGACTCACTATAGGG CGATGCTGAAGGTTGGCTAC-3′
Dalotia bubblegum (*bgm*) R	This Study; Integrated DNA Technologies, Inc	5′-TAATACGACTCACTATAG GGCAATTTCAATGTGGGCCCCA-3′
Dalotia copper-transporting ATPase 1 (*ATP7*) F	This Study; Integrated DNA Technologies, Inc	5′-TAATACGACTCACTATAGGTGA CAACGCAGGATATCCCTCCGG-3′
Dalotia copper-transporting ATPase 1 (*ATP7*) R	This Study; Integrated DNA Technologies, Inc	5′-TAATACGACTCACTATAGGGCT TCTGGTTTCACAGGATCCGCC-3′
Dalotia β-glucosidase (*BGLU*) F	This Study; Integrated DNA Technologies, Inc	5′-TAATACGACTCACTATAGGGCG TGCGCGTGTTGATTACGTC- 3′
Dalotia β-glucosidase (*BGLU*) R	This Study; Integrated DNA Technologies, Inc	5′- TAATACGACTCACTATAGGTGC AGTAACGCGAACGCCATCA-3′
Oligos of *CYP4G* genomic flanking sequence of Hypocyphtines	See Data S1M	N/A
Software and algorithms		
FastQC v0.11.8	Andrews[125]	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
kmergenie v.1.7048	Chikhi and Medvedev[126]	http://kmergenie.bx.psu.edu/
GenomeScope v1.0	Vurture et al.[127]	http://genomescope.org/
covEST v0.5.6	Hozza et al.[128]	https://github.com/mhozza/covest
findGSE v0.1.0	Sun et al.[129]	https://github.com/schneebergerlab/findGSE
jellyfish v2.2.10	Marçais and Kingsford[130]	https://www.genome.umd.edu/jellyfish.html
Smudgeplot	Ranallo-Benavidez et al.[131]	https://github.com/KamilSJaron/smudgeplot
MEGAHIT v1.1.3	Li et al.[132]	https://github.com/voutcn/megahit
Blobtools v1.0	Laetsch and Blaxter[133]	https://github.com/DRL/blobtools
Redundans v0.14a	Pryszcz and Gabaldón[134]	https://github.com/Gabaldonlab/redundans
GapCloser v1.12	Luo et al.[135]	https://sourceforge.net/projects/soapdenovo2/files/GapCloser/src/r6/
SOAPdenovo2-fusion v2.04	Luo et al.[135]	https://github.com/aquaskyline/SOAPdenovo2
SSPACE-LongRead v1.1	Boetzer and Pirovano[136]	https://github.com/Runsheng/sspace_longread
WTDBG2 v2.3	Ruan and Li[137]	https://github.com/ruanjue/wtdbg2/releases/tag/v2.3
canu v1.8	Koren et al.[138]	https://github.com/marbl/canu/releases/tag/v1.8
quickmerge v0.3	Chakraborty et al.[139]	https://github.com/mahulchak/quickmerge
LR_Gapcloser v1.1	Xu et al.[140]	https://github.com/CAFS-bioinformatics/LR_Gapcloser
Pilon v1.23	Walker et al.[141]	https://github.com/broadinstitute/pilon/releases/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
racon v1.3.3	Vaser et al.[142]	https://github.com/isovic/racon/releases/tag/1.3.3
Purge Haplotigs	Roach et al.[143]	https://bitbucket.org/mroachawri/purge_haplotigs/src/master/
Bionano Solve v3.7.1	Bionano	https://bionano.com/software-downloads/
HybridScaffold v11657	Bionano	https://bionano.com/software-downloads/
RefAligner v12432	Bionano	https://bionano.com/software-downloads/
cutadapt v1.18	Martin[144]	https://cutadapt.readthedocs.io/en/v1.18/installation.html
bowtie2 v2.3.4.1	Langmead and Salzberg[145]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
samtools v1.8	Danecek et al.[146]	https://www.htslib.org/download/
HiCAssembler v1.1.1	Renschler et al.[147]	https://pypi.org/project/HiCAssembler/
minimap2 v2.15	Li[148]	https://github.com/lh3/minimap2/releases/tag/v2.15
BUSCO v4.1.1	Manni et al.[55]	https://busco.ezlab.org/
RepeatModeler v 1.0.11	Smit and Hubley[149]	https://github.com/Dfam-consortium/RepeatModeler/releases/tag/open-1.0.11
MITE tracker	Crescente et al.[150]	https://github.com/INTABiotechMJ/MITE-Tracker
vsearch v 2.7.1	Rognes et al.[151]	https://github.com/torognes/vsearch/releases/tag/v2.7.1
RepeatMasker v 4.07	Smit, A., Hubley, R., and Green, P. RepeatMasker Open-4.0.	https://www.repeatmasker.org/
dnaPipeTE v1.3.1	Goubert et al.[152]	https://github.com/clemgoub/dnaPipeTE
RepeatExplorer2	Novák et al.[153]	http://repeatexplorer.org/?page_id=818
seqtk v1.3	N/A	https://github.com/lh3/seqtk
bedtools v2.26.0	Quinlan and Hall[154]	https://bedtools.readthedocs.io/en/latest/
RepeatProfiler v1.1	Negm et al.[155]	https://github.com/johnssproul/RepeatProfiler
TideHunter v1.2.2	Gao et al.[156]	https://github.com/Xinglab/TideHunter
FlexiDot v1.06	Seibt et al.[157]	https://github.com/molbio-dresden/flexidot
VectorBuilder	N/A	https://en.vectorbuilder.com/tool/dna-secondary-structure.html
GeneMark-ES v4.33	Lomsadze et al.[158]	http://topaz.gatech.edu/GeneMark/
BRAKER v2.1.2	Brůna et al.[159]	http://topaz.gatech.edu/GeneMark/braker.html
PASA v 2.3.3	Haas et al.[160]	https://github.com/PASApipeline/PASApipeline/releases/tag/pasa-v2.3.3
exonerate	Slater and Birney[161]	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
GeMoMA v1.6.1	Keilwagen et al.[162]	https://www.jstacs.de/index.php/GeMoMa
STAR v2.6.1	Dobin et al.[163]	https://github.com/alexdobin/STAR/releases/tag/2.6.1a
TRINITY v2.5.1	Haas et al.[164]	https://github.com/trinityrnaseq/trinityrnaseq/releases/tag/Trinity-v2.5.1
GMAP v 2017-11-15	Wu and Watanabe[165]	http://research-pub.gene.com/gmap/
blat	Kent[166]	https://github.com/djhshih/blat
EVidenceModeler	Haas et al.[160]	https://github.com/EVidenceModeler/EVidenceModeler
Liftoff v1.6.1	Shumate and Salzberg[167]	https://github.com/agshumate/Liftoff
cath-tools v 0.16.2	Taylor and Christine[168]	https://github.com/UCLOrengoGroup/cath-tools

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
eggNOG emapper v2.1.5	Huerta-Cepas et al.[169]	https://github.com/eggnogdb/eggnog-mapper/releases/tag/2.1.5
cdHIT v4.8.1	Fu et al.[170]	https://github.com/weizhongli/cdhit/releases/tag/V4.8.1
OrthoFinder v2.5.2	Emms and Kelly[171]	https://github.com/davidemms/OrthoFinder/releases/tag/2.5.2
trimAl v1.4.1	Capella-Gutiérrez et al.[172]	https://github.com/inab/trimal
FastTree2 v2.1.11	Price et al.[173]	http://www.microbesonline.org/fasttree/
PhyloPyPruner v1.2.4	N/A	https://pypi.org/project/phylopypruner/
MARE_v0.1.2-rc	Misof et al.[174]	https://bonn.leibniz-lib.de/en/research/research-centres-and-groups/mare
ModelFinder	Kalyaanamoorthy et al.[175]	http://www.iqtree.org/
IQ-TREE v2.2.0-beta	Minh et al.[176]; Hoang et al.[177]	http://www.iqtree.org/
ASTRAL v5.6.3	Zhang et al.[178]	https://github.com/smirarab/ASTRAL/releases/tag/v5.6.3
MCMCtree	Yang[179]	https://github.com/abacus-gene/paml
MCMCtreeR	Puttick[180]	https://github.com/PuttickMacroevolution/MCMCtreeR
Augustus webserver	Stanke et al.[181]	https://bioinf.uni-greifswald.de/augustus/
augustus v3.2.3	Stanke et al.[181]	https://github.com/Gaius-Augustus/Augustus
shoot.bio	Emms and Kelly[182]	https://shoot.bio/
mafft v7.505	Katoh and Standley[183]	https://mafft.cbrc.jp/alignment/software/
HyPhy package v2.5.38	Pond et al.[184]	https://github.com/veg/hyphy
CODEML in the ete3 v3.1.2	Huerta-Cepas et al.[185]	http://etetoolkit.org/
tranalign v6.6.0.0	Rice et al.[186]	https://github.com/kimrutherford/EMBOSS
Tool to infer Orthologs from Genome Alignments	Kirilenko et al.[187]	https://github.com/hillerlab/TOGA
lastz	Harris[188]	https://github.com/lastz/lastz
snpEff v5.0e	Cingolani et al.[189]	https://pcingola.github.io/SnpEff/
bwa v0.1.17	Li and Durbin[190]	https://github.com/lh3/bwa
GATK	Van der Auwera and O'Connor[191]	https://gatk.broadinstitute.org/hc/en-us
bcftools v1.8	Danecek et al.[146]	https://www.htslib.org/doc/1.8/bcftools.html
MUMmer package v 3.23	Kurtz et al.[192]	https://github.com/chienchi/MUMmer
DAGchainer	Haas et al.[193]	https://dagchainer.sourceforge.net/
R package ape v5.6-2	Paradis et al.[194]	https://cran.r-project.org/web/packages/ape/index.html
IMOD software package	Kremer et al.[195]; Mastronarde[196]; Mastronarde and Held[197]	https://bio3d.colorado.edu/imod/
featureCounts v2.0.0	Liao et al.[198]	https://subread.sourceforge.net/featureCounts.html
DESeq2 v1.30.1	Love et al.[199]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
clusterProfiler v3.18.1	Yu et al.[200]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package pheatmap v1.012	Kolde and Kolde[201]	https://cran.r-project.org/web/packages/pheatmap/index.html
R package chisq.posthoc.test v0.1.2	Ebbert[202]	https://cran.r-project.org/web/packages/chisq.posthoc.test/index.html
variancePartition v1.26.0	Hoffman and Schadt[203]	https://www.bioconductor.org/packages/release/bioc/html/variancePartition.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
sva v3.44.0	Zhang et al.[204]	https://bioconductor.org/packages/release/bioc/html/sva.html
AnnotationForge v1.38.0	Carlson and Pages[205]	https://bioconductor.org/packages/release/bioc/html/AnnotationForge.html
R v4.2.1	R Core Team[206]	https://www.r-project.org/
PAL2NAL	Suyama et al.[207]	https://www.bork.embl.de/pal2nal/
OrthoSNAP	Steenwyk et al.[208]	https://github.com/JLSteenwyk/orthosnap
Data analysis and scripts	This study	https://github.com/Parker-Lab-Caltech/Genomic_and_Cellular_Biosynthetic_Innovation_in_Rove_Beetles

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

Plasmids, dsRNAs and enzymes generated for this study are available via request from the lead contact.

Data and code availability

- Sequence reads related to this manuscript have been deposited in the NCBI Sequence Read Archive (SRA) database under the accession numbers listed in the key resources table. New genome assemblies from this study have been deposited in the NCBI GenBank database, with accession numbers listed in the key resources table. Genome assemblies from other studies were downloaded from the NCBI Reference Sequence (RefSeq) database (accessions listed in key resources table). All other data were uploaded to CaltechData (see key resources table for listed DOIs) and are available as of the date of publication.
- Code for the genome assembly, repeat and gene prediction, phylogenomic and phylogenetic tree construction, selection tests, inactivating mutation identification and other analyses has been deposited on GitHub (https://github.com/Parker-Lab-Caltech/Genomic_and_Cellular_Biosynthetic_Innovation_in_Rove_Beetles) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Rove beetles

The *Dalotia coriara* beetles used in this study for genome sequencing, cell-type-specific RNAseq, bulk RNAseq, and all experiments, were from a laboratory-reared stock, GEN7, which has been maintained continuously in the Parker lab. This stock originated from Applied Bionomics (Canada) but was partially isogenized in the Parker lab by seven generations of single pair sibling crosses. GEN7 *Dalotia* were maintained in 10-L plastic food containers inside a climate room set to 25°C and 65% humidity, with an approximate 10HL-14HD light-dark cycle. Rearing containers were prepared with a 1″ layer of slightly damp coconut fiber substrate. Beetles were fed every 2 days with a 50-50 mixture of finely ground oatmeal and poultry pellets.

The *Aleochara* sp. 3 beetles used in this study for genome sequencing, cell-type-specific RNAseq, bulk RNAseq, and all experiments were collected from a rat cadaver trap placed behind the back fence of the corresponding author's home in South Pasadena, CA (see Data S1C for collection data).

The *Liometoxenus newtonarum* beetles used in this study for genome sequencing, cell-type-specific RNAseq, bulk RNAseq, and all experiments, were collected from *Liometopum occidentale* ant colonies in the Angeles National Forest, CA (see Data S1C for collection data).

The *Oligota* and *Holobus* beetles used in this study for genome sequencing, bulk RNAseq, and all experiments, were collected from sifted leaf litter in the Angeles National Forest, CA (see Data S1C for collection data).

The *Cypha* beetles used in this study for genome sequencing and all experiments, were collected from Elsham Parish, UK, by Charlie Barnes (see Data S1C for collection data).

Other rove beetle taxa used to produce draft assemblies were either collected by the authors or were obtained as gifts from other entomologists (see Data S1C for collecting localities and names of collectors).

Sample size and inclusion/exclusion criteria

For Illumina genome sequencing, a single male beetle was used for all species except in the cases of *Oligota*, *Cypha* and *Holobus*, where multiple beetles were pooled due to the minute size of these insects and the low DNA yield per specimen. We used males for genome sequencing to ensure coverage of all autosomes and sex chromosomes.

For *Dalotia coriara* SPRITE, Bionano optical mapping, bulk RNASeq of tissue types and sexes, multiple beetles were pooled to enhance yield. For all datasets aside from female transcriptomes, male beetles were used to ensure coverage of all autosomes and sex chromosomes and for consistency with the Illumina genomic data.

For cell-type specific transcriptome sequencing, BQ cells, solvent cells and control tissue (tergite 6) were collected from multiple individuals, each individual yielding a replicate, of which we obtained $n \geq 5$ for each cell/tissue type per species. For all species, male beetles were exclusively used to enable us to control for possible transcriptional variation arising from sex differences.

METHOD DETAILS

DNA extraction and short and long-read sequencing

Dalotia were collected from an inbred population (original source: Applied Bionomics, Canada) reared in the lab as described previously.[119] Other taxa were collected from various locations or donated to this study (see Data S1C). For Illumina sequencing, DNA was isolated from a single specimen, with the exception of *Cypha longicornis*, *Holobus* sp. and *Oligota* sp. with two, five and seven specimens, respectively. We used either a non-destructive extraction method described by Maruyama and Parker[44] in which the whole specimen is placed in DNA extraction buffer for two days, or a complete tissue homogenization with the Qiagen DNeasy Blood and Tissue extraction kit (Qiagen, Germany) following the manufacturer's protocol with slight modifications as follows. Tissue was homogenized in the ATL lysis-proteinase K solution and incubated for 4 h or overnight at 56°C. The tissue solution was incubated in RNaseA (Qiagen, Germany) for 2 min followed by the manufacturer's protocol. Two rounds of DNA elution were performed with 100 μL warmed elution buffer (50°C) each round. For the non-destructive protocol, specimens were air dried and incubated in SDS-based DNA extraction buffer (3 mM CaCl2, 2%sodium dodecyl sulfate (SDS), 40 mM dithiotreitol (DTT), 250 mg/mL proteinase K, 100 mM Tris buffer pH 8 and 100 mM NaCl)[209] for two days at 55°C with periodic agitation. DNA was extracted from the solution using an equal volume of 25:24:1 phenol/chloroform/isoamyl alcohol solution (Sigma Aldrich), followed by a back-extraction on the organic phase using equal volume of 50mM Tris and 15 mM NaCl, and a final chloroform only extraction on the combined extracts. The DNA was precipitated in 100% ethanol with 1/10th the volume of 3M NaOAc and 1 μL of Glycogen (Thermo Fisher Scientific) and washed with 70% ethanol.

The DNA was concentrated using the Monarch PCR and DNA Cleanup kit (New England Biolabs, MA) with warmed elution buffer. DNA quantity was assessed using the Qubit High Sensitivity dsDNA kit (Thermo Scientific, MA) and DNA integrity was assessed visually with gel electrophoresis. To complement field identifications, we also amplified fragments of *cytochrome c oxidase subunit 1* and *18S* rRNA for each specimen according to Maruyama and Parker (2017) with the Advantage 2 polymerase mix (Takara, Kusatsu, Shiga, Japan). PCR products were purified using ExoSAP-IT (ThermoFisher, MA) and sequenced by Laragen (Culver City, CA). Illumina paired-end sequencing libraries were prepared using the Illumina TruSeq DNA (Illumina, CA) or NEBNext Ultra FS DNA library kits (paired-end 150bp reads, average insert size 155 ± 105 bp, New England Biolabs, MA)) following the manufacturer's protocol, quantified with Agilent Bioanalyzer High Sensitivity DNA kit (Agilent Technologies, CA) and sequenced on various Illumina platforms by Iridian Genomics, Macrogen (now Psomagen), Fulgent Genetics, Genewiz, and the Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech (Data S1L). For *Dalotia*, two rounds of MinION Nanopore vR9 sequencing libraries (Oxford Nanopore Technologies, UK) were prepared using genomic DNA extracted from approximately 25 male beetles using the Qiagen MagAttract HMW DNA Kit (Qiagen, Germany) and run on MinION flow cells at the Millard and Muriel Jacobs Genetics and Genomics Laboratory, Caltech.

Bionano optical mapping

Optical maps were generated on the Bionano Genomics Saphyr system from ~3 μg of ultra-high molecular weight genomic DNA extracted from 100 2nd and 3rd instar *Dalotia* larvae by HistoGenetics (Ossining, NY). The genomic DNA was fluorescently labeled with restriction enzyme DLE-1 (motif CTTAAG) with an average labeling density of 13 per 100 kbp. Total amount of labeled DNA was 755.67 Gbp. The raw Bionano data is available at CaltechDATA: https://doi.org/10.22002/1914a-m9460.

SPRITE

For the Split-Pool Recognition of Interactions by Tag Extension (SPRITE) protocol, 92 male *Dalotia* were prepared as described in Quinodoz et al.[54] with some modifications. Beetles were macerated with a glass dounce in 8 mL of 2 mM disuccinimidyl glutarate cross-linking solution at room temperature and rocked gently for 45 min. The cell suspension was pelleted by centrifugation for 8 min at 2500 xg at room temperature, rinsed in PBS and re-pelleted. A 3% paraformaldehyde solution in PBS was added and rocked gently at room temperature for exactly 10 min followed by the addition of 2.5 M glycine solution at room temperature for 5 min to quench the crosslinking reaction. Cells were pelleted by centrifugation at 4°C for 4 min at 2500 x g. The pellet was washed in cold 1x PBS and 0.5% BSA two times, aliquoted, flash frozen in liquid nitrogen and stored at −80°C. Cells pellets were thawed on ice and then lysed using buffers A, B and C in the SPRITE protocol[124] with buffer exchanges following centrifugation at 2500 xg for 8 min. The lysed cells were sonicated at 4°C for 1 min (0.7s on, 3.3s off) with a chilled Branson needle tip sonicator. DNA fragmentation of lysate was performed with the addition of 3 μL of Turbo DNase (Thermo Fisher, CA) to 5 μL of lysate, 2 μL of 10X SPRITE DNase Buffer, and 5 μL of water at 37°C for 17 min to obtain a fragment size distribution between 50 and 1000 bp. The cross-links were then reversed and the remainder of the protocol was followed as previously described. The distribution of cluster sizes and ligation efficiency was checked with an Illumina MiSeq run in house prior to shipping the twenty paired-end libraries for sequencing on the Illumina HiSeqX by Fulgent Genetics.

Illumina genome assemblies

Read quality for each taxon was assessed using FastQC v0.11.8.[125] Illumina adapters, low-quality nucleotide bases (phred score below 15) from the 3′ and 5′ ends and reads shorter than 50 bp were removed using cutadapt v1.18.[144] From the filtered reads, *in silico* genome size estimates were calculated using kmergenie v.1.7048[126] GenomeScope v1.0,[127] covEST v0.5.6,[128] and findGSE v0.1.0.[129] The latter three required a *k-mer* histogram computed by jellyfish v2.2.10[130] with *k-mer* size of 21. The *in silico* estimates were compared to flow cytometry estimates for *Dalotia* (n = 13 female and *n* = 14 male adult heads, and 3rd and 1st stage instars) performed by Dr. J. Spencer Johnston at Texas A&M University. Samples were run on a Beckman Coulter Cytoflex flow cytometer against both *Drosophila melanogaster* (1C = 175 Mbp) and *Drosophila virilis* (1C = 328 Mbp) standards as described in Johnston et al. 2019.[210] The ploidy level for each taxon was inferred using Smudgeplot[131] that calculates the coverage of heterozygous *k-mer* pairs from the short-read sequences. A preliminary assembly was constructed from the filtered, adapter-trimmed reads using MEGAHIT v1.1.3[132] with multiple *k-mer* sizes (–k-list = 21, 29, 39, 59, 79, 99, 119). Assembled contigs identified as bacterial contaminants with low GC content, high coverage and blast matches to the nr database (downloaded February 2019, e-value 1e-25) were removed using Blobtools v1.0.[133] For all the genome assemblies, except *Dalotia* described below, the filtered contigs were assembled into scaffolds with three iterations of the Redundans v0.14a[134] reference-based pipeline using the *Dalotia* hybrid assembly (v1) as a reference (–iters 3, –limit 0.5, –nogapclosing). Scaffolds smaller than 1 kb were removed and gaps were filled using GapCloser v1.12.[135]

Dalotia coriaria genome assembly

The *Dalotia* genome was first assembled using a hybrid approach with short and long reads (Figure S1A). Illumina reads were processed and assembled as described above until scaffolding. We removed 1,503 assembled bacterial contigs and 701 scaffolds smaller than 1000 bp prior to short-read scaffolding. Scaffolding was performed using SOAPdenovo2-fusion v2.04[135] with a *k-mer* size of 75 optimized around the “best” k predicted by kmergenie. This was followed by long-read scaffolding with SSPACE-LongRead v1.1[136] using uncorrected Nanopore reads (*n* = 4,150,648) and optimized parameters reported by Karlsson et al.[211] Separately, a long-read assembly was constructed with WTDBG2 v2.3[137] using corrected Nanopore reads (*n* = 848,141) from the *correct* step in canu v1.8.[138] We abandoned using canu beyond this step due to the runtime exceeding one month. The two genome assemblies (hybrid and long-read only) were merged using quickmerge v0.3[139] (-hco 5.0 -c 1.5 -L 800000 -mL 10000) where the WTDBG2 assembly acted as the reference for whole genome alignment with nucmer.[192] The merged hybrid assembly (*Dcor* v1, Data S1A) was polished twice using racon v1.3.3,[142] gap-filled using LR_Gapcloser v1.1.[140] and finished with two additional rounds of short-read polishing using Pilon v1.23.[141] We removed 16.8 Mb of allelic scaffold copies identified by Purge Haplotigs[143] based on both long-read (-l 15 -m 70 -h 100) and short-read (-l 8 -m 51 -h 140) coverage resulting in the *Dcor* v1 assembly.

Consensus optical maps were generated *de novo* using Bionano Solve v3.7.1 and used to reorient and correct mis-assemblies of the *Dcor* v1 assembly using HybridScaffold v11657 (Data S1A). Because only a third of the optical maps aligned to the *Dcor* v1 assembly, we aligned the optical maps to preliminary assemblies and raw reads with read length of 10kb or longer using RefAligner v12432 with default settings to calculate the proportion of contigs or reads not contained within the assembly. Assembly gaps were filled in this new assembly, *Dcor* v2, using LR_Gapcloser v1.1 with uncorrected Nanopore reads.

To get the assembly to the chromosome scale, the SPRITE fastq reads were processed by trimming the adapters using cutadapt v1.18 and identifying reads with five barcode tags using *BarcodeIdentification.jar* and *get_full_barcodes.py* scripts of SPRITE protocol. Complete reads were mapped to the *Dcor* v2 assembly with bowtie2 v2.3.4.1,[145] filtered for mapping quality (-bq 20) and primary mapping (-F 256) using samtools v1.8[146] and grouped into clusters using the *get_clusters.py* script from the SPRITE protocol. Clusters belonging to size classes 2 to 100 were first converted into the cool matrix format using *make_sprite_cooler.sh* script and then converted to the h5 format using hicexplorer v2.1.1.[212] Matrix bin sizes were merged using *hicMergeMatrixBins* (-nb 30) and corrected using *hicCorrectMatrix* (–filterThreshold −2 2) to remove low and high coverage bins. The matrix was then used to orient and scaffold the *Dcor* v2 assembly using HiCAssembler v1.1.1[147] with coordinates of misassemblies identified using the *plotScaffoldInteractive* tool provided (–min_scaffold_length 200000 –bin_size 10000 –misassembly_zscore_threshold −1.0 –num_iterations 4). Pseudochromosomes 1 and 5 were manually split at low contact density regions and renamed using the bedtools “getfasta” tool.[154] The assembly was then gap-filled using LR_Gapcloser v1.1.[140] and polished using Pilon v1.23.[141]

To identify sequences that were not incorporated in the chromosome-resolved assembly, the preliminary assemblies from SSPACE-LongRead and WTDBG2 (both corrected and uncorrected versions, Figure S1A) were mapped back to the SPRITE assembly with minimap2 v2.15 full genome alignment setting (-ax asm5).[148] Unmapped scaffolds/contigs were extracted using samtools v1.8 utilities *view* and *fasta,* filtered using Purge Haplotigs with short-read coverage (-l 20 -m 51 -h 140) and then sequences shorter than 1000 bp were removed. The remaining contigs were combined with the SPRITE assembly for the final assembly version, *Dcor* v3. Genome completeness of *Dcor* v3 and the other genome assemblies used in this study was assessed using BUSCO v4.1.1 with the Arthropoda odb10 orthologous gene set (n = 1013) curated from 90 species.[55]

Repeat identification and masking

To predict repeat content of the genome assemblies, we used a reference-based and a read-based approach. For the assembly-based predictions, we used methods described by Brückner et al.[213] Species-specific libraries were constructed with RepeatModeler v 1.0.11[149] and MITE tracker.[150] Each library was filtered for genuine proteins based on significant blast homology (e-value 1e-5) to a local database of beetle proteins (*Agrilus planipennis*, *Anoplophora glabripennis*, *Aethina tumida*, *Dendroctonus ponderosae*, *Leptinotarsa decemlineata*, *Nicrophorus vespilloides*, *Onthophagus taurus*, and *Tribolium castaneum*; see Data S1C for accessions). Blast reports were manually screened to remove non-repeat hits. Repeats without classification but blast homology to known TEs in the beetle protein database were retained whereas those with no blast homology were removed.[214] The remaining repeat families were combined with the Arthropoda sequences in RepBase and clustered using vsearch v 2.7.1 (–iddef 1 –id 0.8 –strand both).[151] For each genome assembly, RepeatMasker v 4.07[215] was used to soft mask repeats using the filtered repeat library. A summary of the masked repeat content was generated using the “buildSummary.pl” script, a utility of RepeatMasker. We also predicted the repeat content of each species using the adapter-trimmed reads with dnaPipeTE v1.3.1,[152] setting a genome coverage of 0.25 based on the predicted k-mer genome size estimates with two rounds of TRINITY assembly. The predicted repeats were filtered as described above by blast searches against the local database of beetle proteins, and reads counts adjusted to calculate the final repeat content.

We explored additional tools to annotate the repeat content of *Dalotia* given that the most abundant repeats lacked annotation from the dnaPipeTE results for the two *Dalotia* samples (WGS1 and WGS2). We used RepeatExplorer 2 v0.3.8.1–466[153] that incorporates additional repeat databases and a satellite identification pipeline. We randomly subsampled two million paired-end reads from *Dalotia* WGS1 and *Dalotia* WGS2 using the “sample” tool in the program seqtk v1.3 (https://github.com/lh3/seqtk). The reads were uploaded to the RepeatExplorer2 Galaxy portal, and we employed the following procedure as described by Novák et al. 2020: within the portal, the reads were pre-processed to remove sequence adapters and low quality bases and then run through the RepeatExplorer2 with almost all default settings except to automatically filter abundant satellite reads. Only 2% of the reads were used in the analysis due to RAM limitations of the Galaxy portal. Nevertheless, 60% of the reads for both samples were assigned to a 147 bp satellite (Dc-Sat1) that matched the abundant repeats of the dnaPipeTE results and was also present in the assembly-based method (“rnd-5_family-549”). To estimate the abundance of the Dc-Sat1 in the *Dcor* v3 assembly, we used bedtools v2.26.0 “intersect” given the genomic location of repeats predicted by RepeatMasker and bed files of the genomic coordinates of exons, introns and intergenic region boundaries. To see if the Dc-Sat1 was shared among the beetles in this study, five million reads were subsampled from each species and screened for the consensus sequence of Dc-Sat1 using RepeatProfiler v1.1[155] with default settings. Lastly, we estimated the Kimura’s distance, or nucleotide sequence divergence, of the Dc-Sat1 with RepeatMasker on a subset of five million reads followed by RepeatMasker utility scripts “buildSummary.pl” and “calcDivergenceFromAlign.pl”. Long minION reads with abundant copies of Dc-Sat1 as determined by TideHunter v1.2.2 using default settings[156] were visualized using FlexiDot v1.06 with a word size of 147.[157] The secondary structure of the Dc-Sat1 satellite was predicted using VectorBuilder (https://en.vectorbuilder.com/tool/dna-secondary-structure.html).

Dalotia gene predictions and annotation

A combination of *ab initio* (GeneMark-ES v4.33[158] and reference-based (BRAKER v2.1.2,[159] PASA v 2.3.3,[160] exonerate[161] and GeMoMA v1.6.1[162]) tools were used for gene prediction in the *Dalotia* assembly versions as previously described.[213] For BRAKER and PASA, diverse transcriptomic datasets (larvae, pupae, male and female antenna, male and female whole body, female brain, and abdominal segments 6 and 7) were mapped to the *Dalotia* genome *Dcor* v3 using STAR v2.6.1.[163] With the resulting alignment file, a genome-guided transcriptome assembly was constructed with TRINITY v2.5.1[216] as described below. The transcriptome assembly constructed from all tissue types and life-stages was then used for gene prediction with PASA run with the Transdecoder option (https://github.com/TransDecoder/TransDecoder), GMAP[165] and blat[166] aligners, and a maximum intron length of 300 kb. To identify homologs of insect genes, we aligned 3,483,422 insect genes from the UniProt database (downloaded March 2019) to the *Dalotia* genome using exonerate v2.4.0, keeping alignment predictions with at least 80% percent coverage.

For the *Dcor* v1 assembly, gene predictions were combined with EVidenceModeler[160] with the following weights: PASA = 10, BRAKER_HiQ = 4, BRAKER = 1, GeneMark = 1, and exonerate = 1. BRAKER_HiQ predictions were given higher weight because they had >90% coverage of the exon boundaries. Gene predictions from *Dcor* v1 were lifted over to subsequent versions using Liftoff v1.6.1[167] with default settings and the polish option. In place of exonerate in later assembly versions, we used the homology-based prediction tool GeMoMa v1.6.1 with gene models from the beetle phylogenetically closest to *Dalotia* with a previously annotated genome, *Nicrophorus* (Staphylinidae: Silphinae; NCBI: GCF_001412225.1), as well as from the beetle with the highest quality, annotated coleopteran genome, *Tribolium* (Tenebrionidae; NCBI: GCF_000002335.3). We combined all predictions with EVidenceModeler with the following weights: GeMoMa = 4, PASA = 4, Liftoff = 4, BRAKER_HiQ = 4, BRAKER = 1 and GeneMark = 1. The predicted genes were searched against the NCBI nr (February 2019), UniProt (February 2019), PFAM (v 32, August 2018), merops (v 12, October 2017) and CAZy (v 7, August 2018) databases. The hmm-based results of PFAM and CAZy were filtered using cath-tools v 0.16.2[168] (https://cath-tools.readthedocs.io/en/; e-value 1e-5) and the blast-based searches were filtered by the top hit (e-value 1e-5 threshold). Predicted genes were also assigned to orthologous groups using eggNOG emapper v2.1.5[169] against the eggNOG 5.0 database. Gene annotation was assigned by the UniProt hit if the e-value < 1e-10 followed by NCBI annotation if the e-value < 1e-10, and then eggNOG annotation if the e-value < 1e-10. If no homology was recovered, then the gene was annotated as "hypothetical protein". The final assembly and associated annotation files can be downloaded at CaltechDATA: https://doi.org/10.22002/62xxb-mak64.

Gene predictions of other genome assemblies

A similar strategy to gene prediction was used for the remaining genome assemblies presented in this study. When transcriptomic data was available (*Holobus* sp., *Drusilla canaliculata*, *Lissagria laeviuscula*, *Aleochara* sp. 3, and *Liometoxenus newtonarum*), both *ab initio* and reference-based tools were used as described above with slight modifications. In addition to *Nicrophorus* and *Tribolium* gene models, gene models from *Dcor* v2 assembly were used for the homology-based predictions with GeMoMa. The respective genome-guided transcriptome assemblies for each species based on available whole body RNAseq read sets were used as the input of PASA and BRAKER and run as described above for *Dalotia*. EvidenceModeler weights were assigned as follows: PASA = 10, BRAKER_HiQ = 4, BRAKER = 1, GeMoMa = 1, and GeneMark = 1. For species where no transcriptomic data was collected, we only used *ab initio* and homology-based predictions. We used an additional *ab initio* tool augustus v3.23[181] that was run with three different configuration files: honeybee, tribolium2012, and species-specific file based on a random set of 200 genes from the BUSCO training set using the etraining tool. To combine the *ab initio* predictions with GeMoMa predictions, EVidenceModeler weights were GeMoMa = 5, species-specific = 1, honeybee = 1, tribolium2012 = 1, and GeneMark = 1. All Illumina-only genome assemblies are available at CaltechDATA: https://doi.org/10.22002/k8sfv-dw648. Predicted genes of *Aleochara* sp., *Holobus* sp. and *L. newtonarum* were assigned annotation through either orthology to *Dalotia* genes from the OrthoFinder2 results or from eggNOG orthology searches when no *Dalotia* ortholog was found.

Phylogenomic tree construction and dating

For the phylogenomic analysis, we included the genome assemblies of 26 Staphylinidae species from this study and nine published genome assemblies of beetle species spanning the suborder Polyphaga (Data S1C). In the case of the published genome assemblies, we removed predicted isoforms with cdHIT v4.8.1[170] if the pairwise protein sequence identities were at least 98% identical (-c 0.98) for at least 30% of the alignment (-aL 0.3 -aS 0.3). Protein-coding sequences for all species were clustered into orthogroups, a group of orthologous genes, with OrthoFinder v2.5.2 (-M msa -S diamond_ultra_sens -A mafft -T fasttree).[171] The 9,971 mafft sequence alignments of orthogroups that had at least 18 taxa present were then trimmed using the gappyout method of trimAl v1.4.1.[172] An approximate maximum likelihood gene tree was constructed for each trimmed alignment with FastTree2 v2.1.11 (-slow –gamma).[173] To reduce the alignments to a strict set of orthologs, we used PhyloPyPruner v1.2.4. (https://gitlab.com/fethalen/phylopypruner) with the following parameters: –min-len 100 –trim-lb 3 –min-support 0.75 –prune MI –min-taxa 28 –mask pdist –trim-divergent 0.75 –min-pdist 0.01 –min-gene-occupancy 0.1 –subclades subclade.txt –root midpoint –outgroup Apla PPYR. The resulting concatenated supermatrix consisted of 1,300,484 amino acid sites with 3,060 gene partitions. To improve the phylogenetic signal, the information content of each partition was calculated using MARE_v0.1.2-rc with default settings, except to ensure all taxa were retained.[174] The optimized supermatrix from MARE contained 1,520 gene partitions (577,200 aligned amino acid sites).

With the reduced and optimized gene partitions, we constructed the species tree using both maximum likelihood and quartet-based coalescent methods. To find the best substitution model, we ran ModelFinder[175] with a subset of protein models (LG, WAG, JTT, Dayhoff, Q.insect) on the gene partitions and examined the top 10% of the partition merging schemes (-rcluster 10).[217] Using the best-scoring partitioning scheme, a maximum likelihood species tree was estimated from the concatenated supermatrix using IQ-TREE v2.2.0-beta[176] with 1,000 ultrafast bootstrap replicates.[177] For the same set of genes, a coalescent species tree analysis was carried out in ASTRAL v5.6.3[178] using gene trees estimated from the pruned alignments in IQ-TREE following model selection by ModelFinder. Topological support is presented as the quartet support, or gene tree conflict around a given node.

To date the species tree, ten conservative fossil calibration points were selected from a literature survey (Data S1D). This set of fossils contained eight calibration points previously reported for the family Staphylindae.[44] The other two calibration points were selected from recent phylogenomic studies on Coleoptera.[15,218,219] These included bounded constraint on the root of the tree, the Crown Polyphaga (237–293 Ma), and lower bound estimate on Crown Chrysomeloidea (122.5 Ma). Divergence time analysis was performed with MCMCtree and CodeML implemented in the PAML v4.9 package[179] on the concatenated supermatrix and maximum likelihood species tree. As part of the approximate likelihood calibration method, we generated a Hessian matrix in CodeML using empirically estimated base frequencies on the protein supermatrix from the LG substitution matrix (lg.dat) with four rate categories. We obtained 200,000 trees with a sampling frequency every 100 iterations after discarding 20,000 trees as burn-in. Default parameters were set as described in McKenna et al. 2019,[15] namely: seqtype = 2, usedata = 2, clock = 2, RootAge = '3.0', model = 0, alpha = 0, ncatG = 5, cleandata = 0, BDparas = 1 1 0.1, kappa_gamma = 6 2, alpha_gamma = 1 1, rgene_gamma = 2 20 1, sigma2_gamma = 1 10 1, finetune = 1: 0.1 0.1 0.1 0.1 0.01 0.05. For all calibration points except the root age, we applied a soft minimum age using a truncated Cauchy distribution with an offset of 0.1, scale parameter of 1 and left tail probability of 2.5%. At the root, we provided a soft joint bound with an error probability of 0.1 on the minimum and maximum age. Convergence of two independent MCMC runs was checked in Tracer v1.7.2.[220] The final species tree was plotted using the R package MCMCtreeR.[180]

Phylogenetic analyses of select gene families

For select orthogroups of interest, we manually refined gene predictions where necessary and constructed gene trees. We manually screened sequences for the presence of start and stop codons and compared the length of each sequence against the length distribution of all sequences within a given orthogroup. If sequences were flagged as partial, we extracted the corresponding scaffold from the genome and attempted to extend the scaffold using the unfiltered megahits assembly of that species. The extended scaffolds were then re-processed through the Augustus webserver using either the *Apis mellifera* or *Tribolium castaneum* configuration files to re-predict coding sequence. In the case of identifying genomic flanking sequence surrounding CYP4G in *Oligota* and *Cypha*, we also confirmed the re-assembled scaffolds with amplification of 1.5–2.5 kb PCR products using adjusted amplification settings for Takara Advantage 2 polymerase followed by whole plasmid sequencing by Primordium Labs (Arcadia, CA). CYP4G primers are available in Data S1M. We added *Drosophila melanogaster* orthologs to each orthogroup using phylogenetic-informed orthology searches with shoot.bio[182] as well as literature searches. We aligned the protein sequences with mafft v7.505.[183] The protein alignment was then trimmed with trimAl v1.4.1 using the *gappyout* method. A maximum likelihood tree was constructed with both the trimmed and untrimmed protein alignments using IQ-TREE v2.2.0-beta with a 1,000 ultrafast bootstrap replicates. The best protein model was selected by ModelFinder with a subset of substitution models (LG, WAG, JTT, Dayhoff, and Q.insect). From the final gene trees, classification of FARs and esterases followed the nomenclatures of Tupec et al.[221] and Oakeshott et al.[222] respectively, using placement of shared *T. castaneum* and *D. melanogaster* sequences in our study. Curated protein and nucleotide sequences used in the phylogenetic analyses and IQ-TREE tree files can be found at CaltechData: https://doi.org/10.22002/cgsw0-9kk67.

Selection tests and inactivating mutations

We performed positive selection tests on gene trees using the adaptive branch-site random effects likelihood method (aBSREL) in the HyPhy package v2.5.38[184,223] and the branch-site models implemented by CodeML in the ete3 v3.1.2 toolkit.[179,185] Both tools used the codon alignment and gene tree as input. Protein alignments were converted into codon alignments with tranalign v6.6.0.0, a tool within the EMBOSS suite.[179,186] For aBSREL, we tested branches using both an exploratory approach across the whole tree and hypothesis approach on select branches of interest (foreground) against the background. A Likelihood Ratio Test was performed on the fit of the full model on each branch against the null model, where no positive selection rate class is allowed on that branch. For CodeML, we tested the branch-site model on select branches and the model fit was compared against the null model with a likelihood ratio test. For branches under selection, the Bayes-Empirical Bayes method identified codons with signatures of positive selection that had a posterior probability threshold ≥ 0.95.

To determine the strength of selection on the core gland genes in lineages with BQ loss, we also performed RELAX tests in the HyPhy package v2.5.38.[224] In order to obtain a single representative sequence from each species for each gland-containing lineage in the gene tree, we extracted gene ids from the multiple sequence alignments of the 549 gene families that contain the core gland genes with a minimum taxon representation of 50% (18 taxa) after re-running PhyloPyPruner. We also used OrthoSnap v1.3.0[208] with a minimum nodal support of 0.75 to recover pruned alignments in cases where more than one core gland gene was found in the same gene family. The protein sequences were re-aligned with MAFFT, converted to codon-alignment with pal2nal v14,[207] and trimmed with trimAl v1.4.1 using the *gappyout* method. Taxa with gaps exceeding 75% of the alignment were identified using the "*get_sequences_gaps_ratio.py*" utility script in trimAL and removed with the "-selectseqs" parameter. This resulted in 469 alignments for the hypocyphtines and 448 for the *Ecitochara*-group. The species tree was then trimmed to match each filtered codon alignment using a custom script. For each test, we compared the test group, either hypocyphtines or *Ecitochara*-group, against the rest of the higher aleocharines (reference group) (Data S5B). The RELAX analysis first estimated a null model by fitting three dN/dS (omega) classes over the entire species tree and then estimated the selection intensity parameter K on the test branches as the alternative model. The alternative model was compared to the null model with a Likelihood Ratio Test and *p*-values were adjusted for each lineage using a false discovery rate correction with a cutoff of 0.05. Results of the selection test are available on CaltechDATA: https://doi.org/10.22002/gz6w6-g5355.

Inactivating mutations were detected using an orthology-based, reference genome alignment method Tool to infer Orthologs from Genome Alignments (TOGA[187]) for the three *Ecitochara*-group taxa against the *Dcor* v3 assembly. To make alignment chain files, each taxon was aligned to the *Dalotia* assembly twice using the utility script "make_chains.py" (https://github.com/hillerlab/make_lastz_chains) with default settings (K = 2400, L = 3000, H = 2000, Y = 9400, default lastz scoring matrix)[188] and University of California Santa Cruz genome browser settings for insect alignments (K = 2200, L = 4000, H = 2000, Y = 3400, HoxD55.q lastz scoring matrix). The chain files were then used as input for TOGA with the "–fragmented-genome" parameter to infer orthologous genes from multiple aligned contigs. To account for sequencing errors and/or sequence divergence, the predicted gene-inactivating mutations (frameshift insertion/deletions, premature stop codons, splice site mutations and deletions of exons or entire genes) from the core biosynthetic differentially expressed orthologs of the solvent and BQ cells (n = 554) were manually inspected with independent gene predictions for each respective taxon and predicted mutations from snpEff v5.0e[189] using a variant call file (VCF) produced by aligning the short reads of each ecitocharine taxon to the *Dalotia* genome assembly with bwa v0.1.17,[190] following the GATK best practices pipeline,[191] and filtering SNPs ('MQ > 40 & INFO/DP < 1200 & QD > 2.0 & FS < 60.0 & MQRankSum > −12.5 & ReadPosRankSum > −8.0 & SOR <3.0) and INDELS (MQ > 40 & INFO/DP < 1200 & QD > 2.0 & FS < 200.0 & ReadPosRankSum > −20.0 & SOR <10.0) with bcftools v1.8.[146] Given the fragmentation of our assemblies from the *Ecitochara*-group taxa (Data S1C), we excluded predicted "loss" genes if the evidence was based solely on missing and/or deleted exons. Mutations were visualized using the “plot_mutations.py” utility script. The results of TOGA and annotated VCF files from snpEff are available on CaltechData: https://data.caltech.edu/records/6xjn1-e3085.

Gene synteny

We compared the gene content and identified the sex chromosomes of the *Dalotia* genome assembly against the chromosome-scale genome assemblies of the outgroup beetles *T. castaneum* (NCBI: GCF_000002335.3) and *P. pyralis* (http://www.fireflybase.org/), and two rove beetles *Ocypus olens* (NCBI: GCA_910593695.1) and *Philonthus cognatus* (NCBI: GCA_932526585.1). Gene synteny was assessed using the “promer” and “show-coords” programs within the MUMmer package v 3.23 with an alignment length of at least 100 amino acids (-L 100) and percent identity of 50% (-I 50) between the reference and target genomes. To identify regions of gene synteny between all pairwise genome comparisons, the all-vs-all blast results from OrthoFinder were used as input for DAGchainer (-M 50 -D 5 -g 1 -A 3 -E 10).[193]

Gland volatile quantification

Beetles were individually submersed in 70 μL hexane (NN), 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 × 0.25 mm ID, df = 0.25 μm) from Phenomenex (Torrance, CA) was used to profile the gland contents: crude sample aliquots (2 μL) were injected into 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: column temperature was set to 40°C with a 1-min 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 min. 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 previously identified and in addition authentic standards were used to construct four-point calibration curves for external standardization and quantification of benzoquinones, esters and alkanes.

Ancestral state reconstruction

We used ancestral state reconstruction to estimate chemical class evolution along the species tree. Each chemical class was treated as a binary, discrete character of either present (1) or absent (0) in a given extant lineage. Extant taxa for which no chemical data has been collected were assigned a value of “-9”. We first applied a maximum likelihood method using an equal rates model with the *ace* command in R package ape v5.6-2.[194] Second, we used the re-rooting method of Yang et al.[225] to estimate marginal states for species with no chemical data implemented in phytools v1.0-3.[226] Probabilities of the state being absent were assigned a value of 0.5 in *Aleochara* sp1, *Falagria* and *Earota* and 0.9 in the *Ecitochara*-group species.

Biochemical tracer experiment in *Liometoxenus*

Wild caught *Liometoxenus* individuals were housed in 10 cm plastic containers with moistened tissue paper for several days with various food sources (sugar water, dead ants and frozen fruit flies) prior to experimentation. Ten beetles were chemically disarmed on CO_2 gas as previously described for *Dalotia*[34] and split into two containers, one with the same food sources and the other where the stable isotope precursor $^{13}C_6$-tyrosine (>99% enrichment, Sigma-Aldrich, MO) was added to each food source. The isotope-labeled and control food was refreshed every three days. Beetles were sacrificed over the course of two weeks for hexane extractions either because their health declined, or the end of the experiment was reached. Hexane extracts were analyzed with a GC-MS as described above. Electron ionization mass spectra of characteristic fragment ions were monitored in single ion mode (SIM) and at 70 eV.

Double-stranded RNA synthesis and knockdown

Double-stranded RNA constructs were prepared as previously described.[34,119] Our target sequences were cloned into a pCR2.1-TOPO vector (Thermo Fisher, CA) using primers with T7 linkers as follows: *very long-chain-fatty-acid-CoA ligase bubblegum* (*bgm*) F: (5′-TAATACGACTCACTATAGGGCGATGCTGAAGGTTGGCTAC-3′) and R: (5′-TAATACGACTCACTATAGGGCAATTTCAATGTGGGCCCCA-3′), *copper-transporting ATPase 1* (*ATP7*) F: (5′-TAATACGACTCACTATAGGTGACAACGCAGGATATCCCTCCGG-3′) and R: (5′-TAATACGACTCACTATAGGGCTTCTGGTTTCACAGGATCCGCC-3′), and *β-glucosidase* (*BGLU*) F: (5′-TAATACGACTCACTATAGGGCGTGCGCGTGTTGATTACGTC- 3′) and R: (5′- TAATACGACTCACTATAGGTGCAGTAACGCGAACGCCATCA-3′). After synthesis, the dsRNA was cleaned using the MEGAclear Transcription Clean-Up kit (Thermo Fisher, CA) and quantified on the NanoDrop (Thermo Fisher, CA). Target and control, a green fluorescent protein, constructs were diluted with DEPC-treated 1x PBS and 1 μL of blue food dye to a final concentration of 2 mg/mL. The constructs were microinjected into third-instar larvae from the laboratory *Dalotia* population according to Parker et al.[119] Following injection, larvae were reared in individual 5 cm Petri dishes on filter paper until eclosion. After eclosion, adult beetles were moved into new Petri dishes and fed frozen fruit flies for ten days, at which point they were used for chemical analysis. Statistical difference of glandular secretions of specific compounds between RNAi-treated and GFP-treated was tested with Wilcoxon signed rank test with a Bonferroni multiple test correction for the various compounds per beetle.

Drosophila toxicity bioassay

We tested the toxicity of the major compounds of the *Holobus* gland secretion on survival of *Drosophila melanogaster* larvae as previously described.[34,70] The major compounds were prepared to mimic natural ratios of the gland secretion: 5% of tridecane, 15% of 2,5-dimethoxybenzaldehyde, and 80% of ethyl linoleate (all Sigma Aldrich, MO) (Figure 6A). Each compound was tested independently along with the mixture of all three compounds in the *Holobus* glandular secretion. We also tested the addition of 2-methyl-1,4-benzoquinone without a solvent (powder form) and with the *Holobus* secretion mixture (28 mg). A mixture of the *Dalotia* gland secretion compounds[34] and 1x PBS were used as the positive and negative controls, respectively. Over two experimental trials, wandering third instar *Drosophila* larvae were submerged in the 1 mL of the various mixtures for ~1 s or dipped in solid BQ powder ($n = 25$ larvae per mixture) and moved to three replicate culture tubes. Survival was scored after 1 h and after eclosion. At 1 h, dead larvae were distinguishable by a change in coloration to black or dark down, or loss of tissue integrity. Differences in survival were tested using an ANOVA with a Tukey *post hoc* test correction in the statistical package R v4.2.1.

Chromosome squashes

The chromosome preparation protocol was modified from Rożek et al.[227] Testes of immobilized *Dalotia* (n = 10) were dissected in 1x PBS under a stereomicroscope. Testes were transferred to a hypotonic KaryoMAX Colcemid Solution (Gibco, NY) at a final concentration of 0.5 μg/mL in 1x PBS for 1 h at room temperature with gentle rocking. The solution was discarded after 2 min centrifugation at 500 xg and replaced with 2 mL of 0.075M KCl for 20 min. Following another round of centrifugation, the testes were transferred to freshly prepared Fix I solution (3:1 absolute 96% ethanol:glacial acetic acid) and left to sit for 30 min at room temperature. The Fix I solution was replaced after 30 min with fresh Fix I and stored at 4°C for up to two years. The remaining fixative solutions (Fix II – 1:1 absolute 96% ethanol:glacial acetic acid and Fix IV – 7:2:1 glacial acetic acid: absolute 96% ethanol:distilled water) were prepared fresh and brought to 32°C when preparing for the squashes. The testes were transferred from Fix I to Fix II and then Fix II to Fix IV, with 30 min incubation intervals in each solution at room temperature. The testes were stored in Fix IV at 4°C overnight for 10–12 h. Fixed testes tissue was then transferred to a clean microscope slide resting on blotting paper. The tissue was macerated quickly using dissecting needles in a few drops of 70% acetic acid. The tissue was squashed between two microscope slides as described in[227] and frozen on dry ice. The final preps were stained with nuclear stain Hoechst 33342 (1:2000), mounted in 25 μL of VectaShield Mounting Media (Vector Laboratories, CA) and imaged using the 100x objective on the Zeiss LSM 880 Confocal Laser Scanning Microscope (Zeiss, Germany).

Electron microscopy and dual-axis tomography

For sample preparation, beetle abdomens were dissected in a fixative comprising 3% glutaraldehyde, 1% paraformaldehyde, 5% sucrose in 0.1 M sodium cacodylate trihydrate. Dissected tissue was then placed in fresh fixative at 4°C. Pre-fixed segments were rinsed with fresh cacodylate buffer and placed individually into brass planchettes (Type A; Ted Pella, Inc., CA) prefilled with 10% Ficoll in cacodylate buffer. Samples were covered with the flat side of a Type-B brass planchette and then ultrarapidly frozen with an HPM-010 high-pressure freezing machine (Bal-Tec/ABRA, Switzerland). The vitrified samples were transferred under liquid nitrogen to cryotubes (Nunc) containing a frozen solution of 2.5% osmium tetroxide, 0.05% uranyl acetate in acetone. Tubes were loaded into an AFS-2 freeze-substitution machine (Leica Microsystems, Vienna) and processed at −90°C for 72 h, warmed over 12 h to −20°C, held at that temperature for 6 h, then warmed to 4°C for 2 h. The fixative was removed, and the samples rinsed 4x with cold acetone, following which they were infiltrated with Epon-Araldite resin (Electron Microscopy Sciences, PA) over 48 h. Samples were flat-embedded between two Teflon-coated glass microscope slides and the resin was polymerized at 60°C for 48 h.

Flat-embedded beetle segments were observed by phase-contrast LM to determine sample quality and specifically locate suitable tergal gland components. These regions were extracted with a microsurgical scalpel, oriented for *en face* (dorsal-to-ventral) sectioning and glued to the tips of plastic microtomy stubs. Semi-thick (170 nm) serial sections were cut with a UC6 ultramicrotome (Leica Microsystems, Vienna) using a diamond knife (Diatome, Ltd. Switzerland). Sections were placed on Formvar-coated copper-rhodium slot grids (Electron Microscopy Sciences, PA) and stained with 3% uranyl acetate and lead citrate. Gold beads (10 nm) were placed on both surfaces of the grid to serve as fiducial markers for subsequent tomographic image alignment. Grids were placed in a dual-axis tomography holder (Model 2040, E.A. Fischione Instruments, PA) and imaged with a Tecnai T12 transmission electron microscope (120 keV) equipped with a 2k x 2k CCD camera (XP1000; Gatan, Inc. Pleasanton CA). Tomographic tilt-series and large-area montaged overviews were acquired automatically using the SerialEM software package.[228] For tomography, samples were tilted +/− 62° and images collected at 1° intervals. The grid was then rotated 90° and a similar series taken about the orthogonal axis. Tomographic data was calculated, analyzed and modeled using the IMOD software package[195–197] on iMac Pro and Mac Studio M1 computers (Apple, Inc., Cupertino, CA).

RNA extraction and transcriptome assemblies

Specimens used for transcriptome sequencing (*Aleochara* sp. 3 male body ($n = 1$), *Dalotia* male antenna (n = approx. 100), female antenna (n = approx. 100), female brain ($n = 1$), larvae (n = approx. 100), pupae (n = approx. 20), male body ($n = 1$), female body ($n = 1$), *Holobus* male body ($n = 5$), and *Liometoxenus* male head ($n = 1$) and body ($n = 1$)) were either extracted live or from flash-frozen material stored at −80°C. Total RNA was extracted from the different species, life stages and tissue types using either the ZYMO *Quick*-RNA Tissue/Insect extraction kit (ZYMO Research, CA) or a combination of Trizol (Life Technologies, CA) and Qiagen RNeasy

Mini kit (Qiagen, Germany) extraction protocol as previously described[229] (see Data S1L). RNA integrity and quantity was assessed with the Nanodrop (Thermo Fisher, CA) and Bioanalyzer High Sensitivity RNA Analysis kit (Agilent Technologies, CA). Paired-end, 150bp sequencing libraries were prepared using the Illumina TruSeq RNA library kit by various companies listed in Data S1L and sequenced on Illumina HiSeq X platform (Illumina, CA).

Transcriptomes used in gene predictions described above were either assembled *de novo* (*Liometoxenus*) or from genome-guided RNAseq read alignments (*Dalotia*, *Holobus* and *Aleochara* sp. 3) with Trinity v2.5.1[164] using the diverse datasets available for each species (Data S1L). For the genome-guided assemblies, adapter-trimmed RNAseq reads were aligned to each respective reference genome using STAR v2.6.1[163] and assembled with the maximum intron length of 10000bp and jaccard clip option in Trinity. Previously published *de novo* assembled transcriptomes of *Drusilla* and *Lissagria,* both construced from male and female whole body RNAseq reads, were also used in gene predictions.[34]

SMART-seq transcriptome sequencing

Microdissection of the specific gland cell types from *Aleochara* and *Liometoxenus* was performed as previously described.[34] This resulted in 3–7 BQ cells, ~1000 solvent cells, or ~1000 control cells from tergite 6 per replicate. Similar to performing microdissections of *Dalotia* tergal gland cell types, contamination from adjacent cells is unlikely in *Aleochara* and *Liometoxenus* due to the spatially discrete nature of BQ and solvent cells in these species. However, due to the size of *Holobus* (Figure S6E), the entire tergite 6 (control) and tergite 7 (gland segment) were dissected in ice-cold DEPC-treated PBS, flash frozen and stored at −80°C until processed. Library preparation was done from either frozen cells or Trizol extracted total RNA (3 out of 4 *Aleochara* control samples) using the NEBNext Single Cell/Low Input RNA Library Prep Kit for Illumina together with NEBNext Multiplex Oligos (New England Biolab) according to the manufacturer's protocol. PCR cycles during the cDNA amplification step varied depending on the sample type and species. For example, in *Aleochara*, cycles ranged from 9 PCR cycles for total RNA input, 14 PCR cycles for solvent cells up to 20 PCR cycles for BQ cells. All *Holobus* preps were held for 14 PCR cycles and all *Liometoxenous* preps were held for 20 PCR cycles. Final library amplification ranged from 8 to 12 PCR cycles depending on the intermediate concentrations of the library during the procedure. 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. The 50bp libraries were sequenced on Illumina HiSeq2500 or NextSeq 2000 with about 20–30 million reads per library by Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech.

Differential expression analysis

SMART-Seq reads were aligned to each respective species genome assembly with STAR v2.6.1[163] and read counts extracted with featureCounts v2.0.0[198] only considering primary alignments (–primary) that mapped to the same chromosome and strand (-C) with a minimum mapping quality of 10 (-Q 10). Genes with fewer than 10 read counts for the minimum group size of a given species and cell type were removed (*Dalotia* $n = 10$, *Aleochara* $n = 4$, *Holobus* $n = 4$, and *Liometoxenus* $n = 5$). Technical variation among samples was estimated for each species separately using variancePartition.[203] In all cases, variation from sources such as library and sequencing batches, different sequencing platforms, and different extraction methods (Data S1L) were lower than between tissue or cell type comparisons or among sample variation (Data S1F and Data S2). No sample outliers were detected when comparing median $\log_2$ normalized read counts for a given cell type or from principal component analysis of the variance stabilized counts for each species.

Differential gene expression was tested for each species using DESeq2 v1.30.1[199] with the design *tissue type* (BQ cell, Solvent cell, or control) + *batch* for cell-specific datasets of *Dalotia*, *Aleochara* and *Liometoxenus* or *segment type* (gland or non-gland) + *batch* for bulk abdominal segment comparisons of *Holobus* and *Dalotia*. Sequencing batch was added for all species except *Aleochara*, which was processed in one sequencing run. Bulk RNAseq reads from *Dalotia* gland and non-gland segments[34] were processed as above with technical replicates collapsed using "collapseReplicates" function in DESeq2. DEGs were identified in each species for each pairwise comparison of cell type or segment type using a Wald test with adjusted p-value ≤ 0.05. DEGs that displayed cell type enriched expression were those with 2-fold higher $\log_2$ expression in one cell type relative to the other gland cell type or control.

To compare expression among species, variance stabilized count matrices of all genes for each species were joined by the OrthoFinder assigned orthologous groups. In cases where multiple orthologs were assigned to the same orthogroup, genes were sorted by their adjusted p-values from the gland cell type against control tests, with the lowest value selected to represent the orthogroup. The combined count matrix was adjusted for expression attributable to each species using the empirical Bayes method "ComBat_Seq" function[204] in the sva R package.[72] A principal component analysis was performed on the transformed data using prcomp function in the R package Stats v3.6.0. An UpSet plot of the ortholog expression by cell type and species was inspired by customized_upset_plots (https://github.com/cxli233/customized_upset_plots). To summarize gene functions, GO and KEGG enrichment test on core BQ cell and solvent cell DEOs were then performed with clusterProfiler v3.18.1[200] using a false discovery rate q-value cutoff of ≤ 0.05 and the simplify function to reduce similarity in GO terms. A custom gene ontology (GO) database was made for *Dalotia* using GO terms assigned from the eggNOG database and Uniprot blast matches with AnnotationForge v1.38.0.[205]

To explore the conservation of abdomninal gene expression programs (GEPs) identified in *Dalotia* from a prior study[34] with other species, *Dalotia* transcripts with high Z score rank to the cuticle cells and ventral fat body and oenocytes GEPs were mapped to the *Dalotia* gene models using GMAP v 2017-11-15.[165] Spearman correlation of GEP expression between *Dalotia* genes and their corresponding *Aleochara* orthologs was performed using cor.test in R. To get qualitative differences between tissue types and life-stages of *Dalotia*, all transcriptome datasets were mapped to the Dcor v3 assembly using STAR v2.6.1[163] and gene counts extracted using featureCounts v2.0.0. Heat maps were generated from normalized variance stabilized counts from DESeq2 and the R package pheatmap.[201] Sex-biased expression was calculated as the difference in library normalized $\log_2$ counts using the *normTransform* function in DESeq2 for the male and female whole-body transcriptomes. Differences were categorized as 2-, 5- and 10-fold higher in one sex over the other per gene and then tabulated by chromosome. Statistical differences in the proportion of biased genes were found using a Pearson's Chi-square test with Bonferroni correction with R package chisq.posthoc.test v0.1.2.[202]

In situ hybridization chain reaction

DNA probe sets were either purchased from Molecular Technologies (Pasadena, CA; https://www.moleculartechnologies.org/) or generated using the "*insitu*_probe_generator" tool (https://github.com/rwnull/insitu_probe_generator) and the pool of oligos was purchased from Integrated DNA Technologies (Coralville, IA) (Data S1N). DNA HCR amplifier, HCR hairpins as well as hybridization, wash and amplification buffers were purchased from Molecular Technologies. The abdominal sections of adult *Oligota* sp., *Aleochara* sp. 3 and *Dalotia* were fixed as previously described.[34] The amplification and detection stages followed published protocols.[230] Probes were initiated with B1-Alexa546, B3-Alexa647 or B4-Alexa488 amplifiers. After amplification and before the final wash steps, Hoechst 33342 (1:2000) to mark nuclei, and Alexa 488- or Alexa 647-Wheat Germ Agglutinin Conjugate (WGA; 1:200) to label cell membranes 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.

In vitro measurement of dmd enzymatic activity

Purified protein of Dmd from *Dalotia, Aleochara*, *Holobus* and *Ecitophya* was prepared as described by Brückner et al.[34] Enzymatic activity of each protein was first tested against a standard substrate, ABTS. The reaction mixture was prepared as 5 mM MES, 0.3 M CuSO4, and 2 mM of ABTS. 2 mM of laccase was added, and the shift in absorption at 420 nm was traced for 10 min. To compare the ability of the four species' Dmd to covert hydroquinones to benzoquinones, the activity of each enzyme on the substrate 2-methyl-1,4-hydroquinone was used as a proxy for conversion of all hydroquinones these beetles produce (1,4-hydroquinone, 2-methyl-1,4-hydroquinone and 2-methoxy-3-methyl-1,4-hydroquinone). For this assay, 2mM of 2-methyl-1,4-hydroquinone was added to vials containing 5 mM MES and 4nM of Dmd protein. After 10 min, the reaction was halted by heating to 60°C. As a control, we performed a reaction in which no enzyme was added, giving an estimate of baseline auto-oxidation of 2-methyl-1,4-hydroquinone under the same reaction conditions. Samples were analyzed with an HP Agilent 1100 High-Performance Liquid Chromatography system G1312A with DAD detector, equipped with an Eclipse XDB C18 5/N column (250 cm × 3 mm, 5 μm) (instrument housed at the the Water and Environment lab at Caltech). 20 μL of sample was passed through gradients of acetonitrile and water, starting from 96% water for 3 min to 20%, then ramped up to 80% for 3 min and held at 80% for 3 min before dropping to 5% for 5.1 min. UV Vis was set to detect a wavelength of 293 nm. Six replicas were prepared for the control and each species' Dmd protein. Synthetic 2-methyl-1,4-benzoquinone was used to quantify the amount of benzoquinone in nanograms. Differences in benzoquinone concentrations were tested using an ANOVA with a Tukey *post hoc* test correction in R.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical tests, including Spearman's Rank Correlation, Pearson's chi-square test, likelihood ratio test, Wald test, Wilcoxon signed rank test, and ANOVA, in this study used for comparison of gene expression programs, sex-biased expression, selection tests, differential gene expression, RNAi knockdowns, toxicity tests and *in vitro* enzymatic activity, are indicated in method details, figures, and figure legends. Multiple test corrections were applied where indicated in the method details using a Bonferroni Correction or False Discovery Rate. For all tests, an alpha level ≤ 0.05 was used to determine significance. All statistical analyses are performed in R v4.2.1[206] or Python.

Supplemental figures

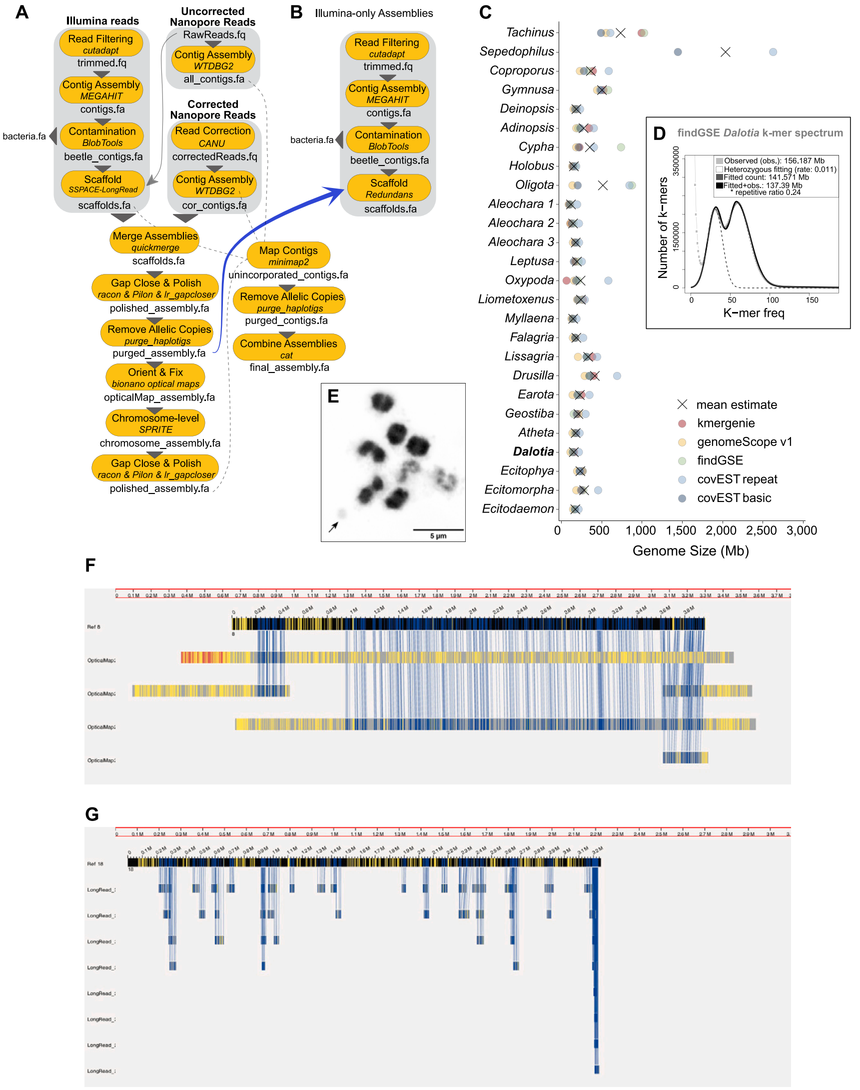


Figure S1. Genome assembly, sizes, and karyotype, related to Figure 2

(A) A schematic of the different bioinformatic steps and datasets (Illumina short reads, Nanopore long reads, BioNano optical maps, and chromatin interaction reads via SPRITE) used to assemble the hybrid genome assembly of *Dalotia coriaria*. Contigs from the preliminary assemblies that did not map to the polished assembly were further filtered to remove putative haplotigs and then combined with the polished assembly for the final genome version (*Dcor* v3).

(B) A schematic of the bioinformatic steps used to assemble the remaining genomes of the samples with only Illumina short-read data.

(C) Estimates of genome size from five *k*-mer based tools (circles: red = kmergenie, yellow = genomeScope, green = findGSE, light blue = covEST repeat, and dark blue = covEST basic; X is the mean estimate).

(D) K-mer frequency histogram of *Dalotia* WGS1 produced with findGSE.

(E) Karyotype of *Dalotia* during mitosis (2*n* = 9+Xyp). The arrow indicates the small Y chromosome.

(F and G) Visualization of Bionano optical map alignments against the *Dcor* v1 assembly and long-reads.

(F) Five optical maps (13, 37, 95, 187, 243) aligned to scaffold ctg4 (ref. 8) from the *Dcor* v1 assembly.

(G) Multiple minION long-reads mapped to optical map 18 (ref. 18) that was not captured by the hybrid assembly process combining optical maps with *Dcor* v1. Aligned labels are dark blue and unaligned labels are yellow along the reference sequence on top (background black) and corresponding query sequences below (background gray) in the genome browser of Bionano Access.

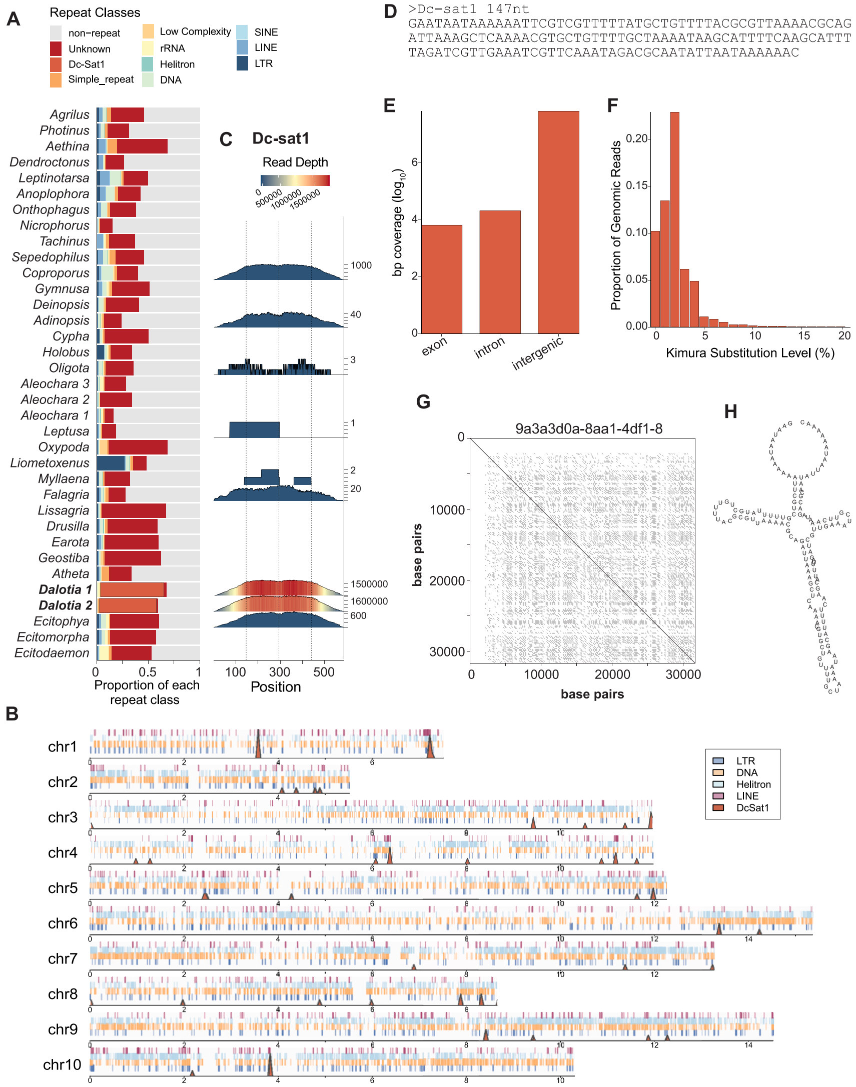


Figure S2. Repetitive content of the aleocharine genomes, related to Figure 2

(A) Predicted proportion of the genome composed of different repeat classes (LTR, LINE, SINE, DNA, Helitron, rRNA, low complexity, simple repeats, satellite *Dc-Sat1*, and unknown) and non-repetitive sequence from 0.25x subsampled short-reads for each respective species using dnaPipeTE v1.3.1. *Dalotia 1* and *Dalotia 2* are two independent short-read assemblies from two separate *Dalotia* specimens.

(B) Distribution of the four most common transposable element classes (LTR, DNA, RC Helitron and LINE) and the *Dc-Sat1* satellite along *Dalotia* chromosomes using the RepeatMasker predictions. DcSat1 is likely underrepresented in this visualization. Other TE classes are largely dispersed throughout the genome with DNA and Helitron TEs showing elevated density along the chromosomes, but not necessarily at the distal (telomeric) or pericentric regions. This lack of association with typically repeat-rich regions is likely due to the underrepresentation of *Dc-Sat1* that makes up the majority of the repeat content (91%) and predicted to be abundant in the centromere and telomeres.

(C) Read depth across four concatenated copies of the 147 bp satellite *Dc-Sat1* from subsampled short-reads from each respective species using RepeatProfiler v1.1. The y axis was adjusted for each species based on maximum read depth.

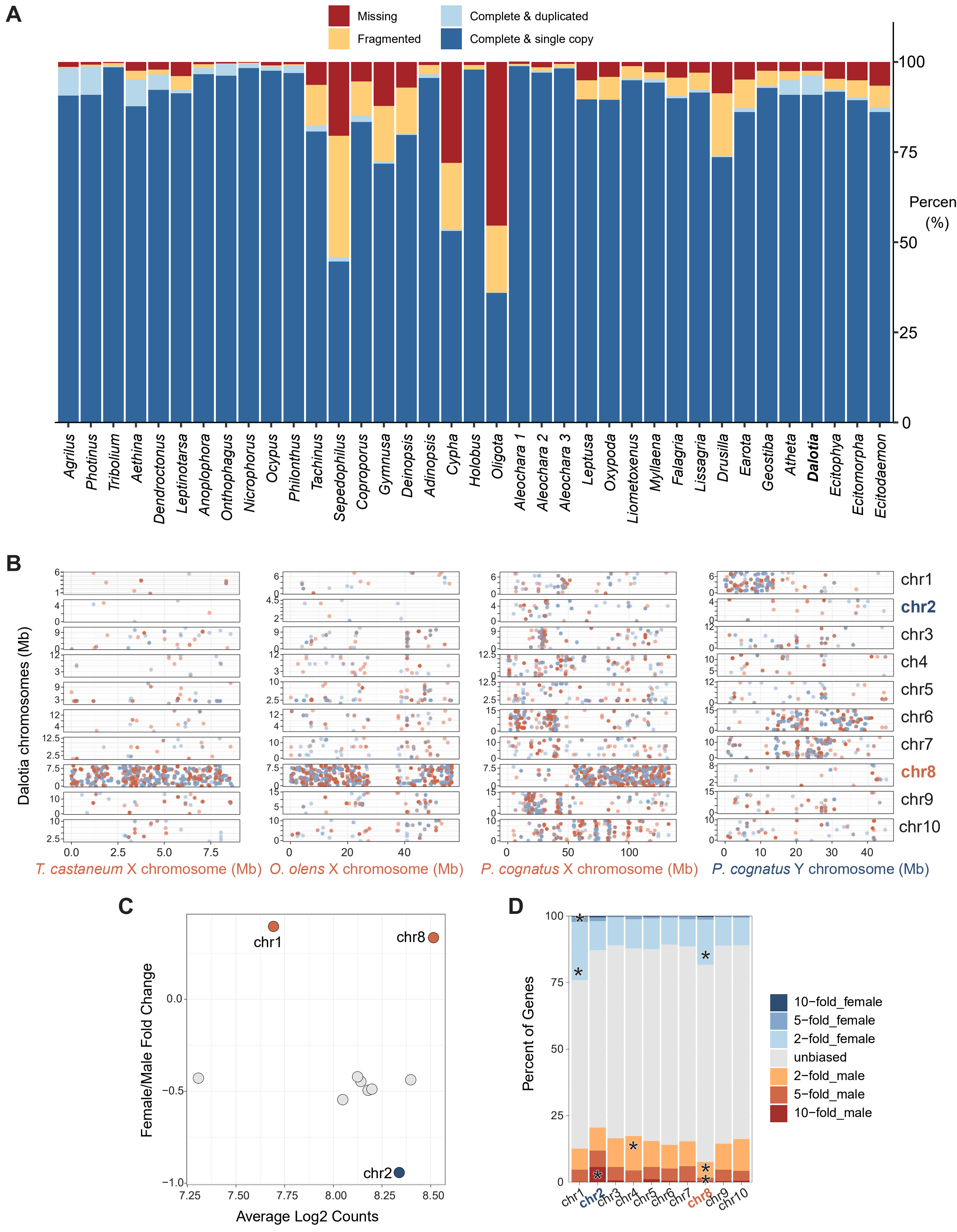
(D) The consensus sequence of *Dc-Sat1* from RepeatExplorer2.

(E) Genome coverage of *Dc-Sat1* in the exons, introns and intergenic regions of *Dcor* v3 assembly.

(F) Estimated proportion of *Dalotia* short-reads with different levels of Kimura substitution, a measure of sequence divergence over time, for *Dc-Sat1*.

(G) Self dot-plot of one example minION read (9a3a3d0a-8aa1-4df1-8) with 35 tandem copies of *Dc-Sat1*.

(H) Predicted secondary structure of *Dc-Sat1*.



(legend on next page)

Figure S3. BUSCO genome completeness assessment for new and previously published beetle genome assemblies, and sex chromosomes of *Dalotia*, related to Figure 2

(A) Percentage of single-copy genes present in the genome assembly of each species using the arthropoda odb10 gene set ($n = 1013$). Dark blue = complete and single copy, light blue = complete and duplicated, orange = fragmented or partial copy, red = missing orthologs.

(B) PROmer amino acid sequence alignment of *Dalotia*'s ten chromosomes against sex chromosomes of *T. castaneum*, *O. olens*, and *P. cognatus*. Alignments were filtered to a minimum length of 200 aa. Each point represents an alignment with percent identity of 50% or higher and colored based on the strand, red is for the negative strand and blue is for the positive strand.

(C) Summarized average $\log_2$ counts for all genes on a given chromosome for both sexes correlated to the fold-change in female to male expression for all genes for a given chromosome. Female-biased expression would have values greater than 0 whereas male-biased expression would be less than 0.

(D) Genes with 2-, 5- and 10-fold difference in normalized $\log_2$ counts between the sexes were tabulated for each chromosome. Categories with significantly over or under representations of genes from Pearson's Chi-square tests adjusted for multiple testing are indicated by asterisk.

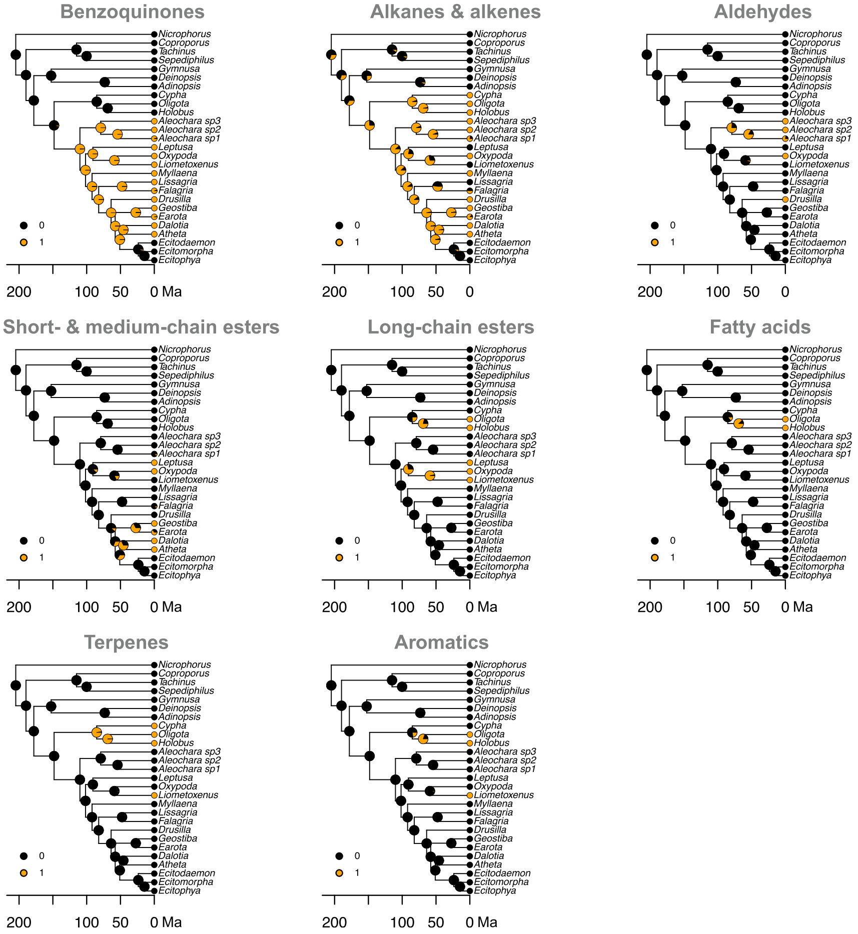


Figure S4. Ancestral state reconstruction of chemical classes found in the tergal gland, related to Figure 3

Pie charts at the nodes represent the maximum likelihood estimates of chemical class evolution along the dated species tree, starting at *Nicrophorus vespilloides*. Each chemical class was marked as present (1 = orange) or absent (0 = black) for extant species from the GC/MS data presented in Figure 3B. If no chemical data were available, we provided a probability of the chemical being absent as 0.5 in *Aleochara* sp1, *Falagria* and *Earota* and 0.9 in the *Ecitochara*-group clade based on morphology and chemical data from their closest sister taxon.

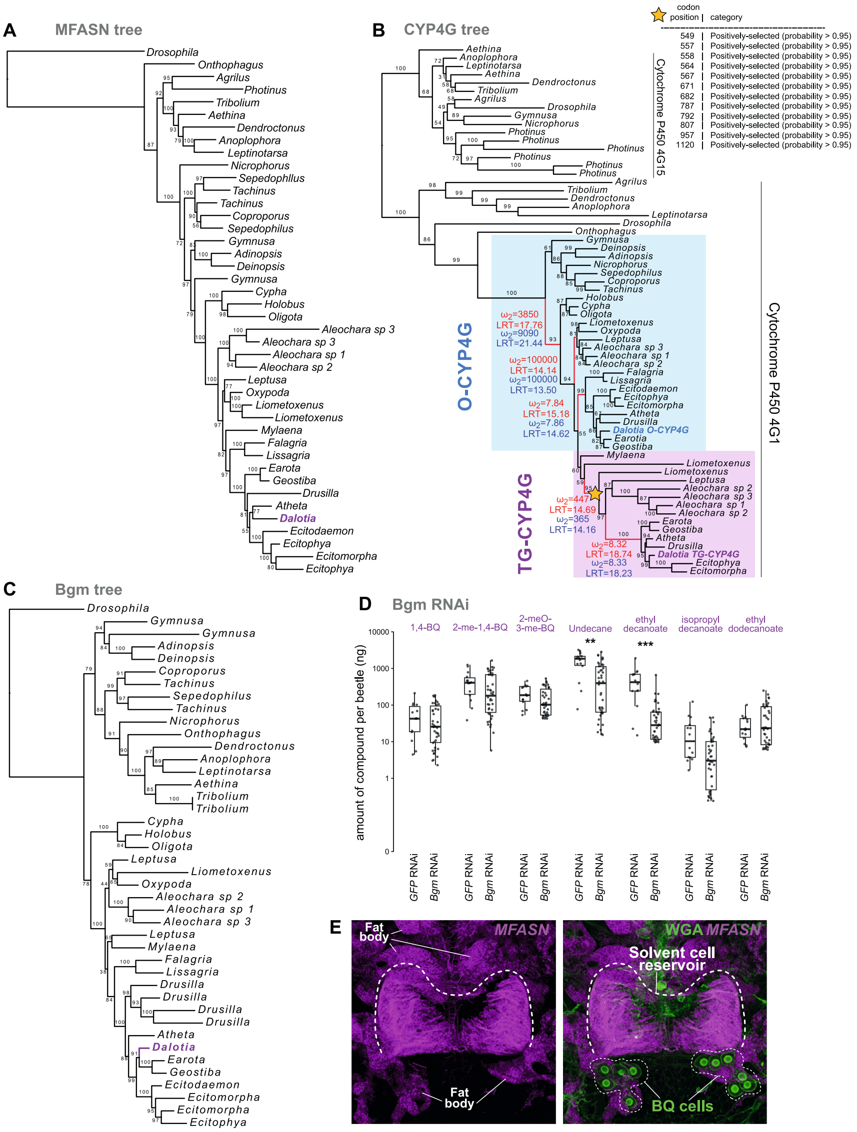
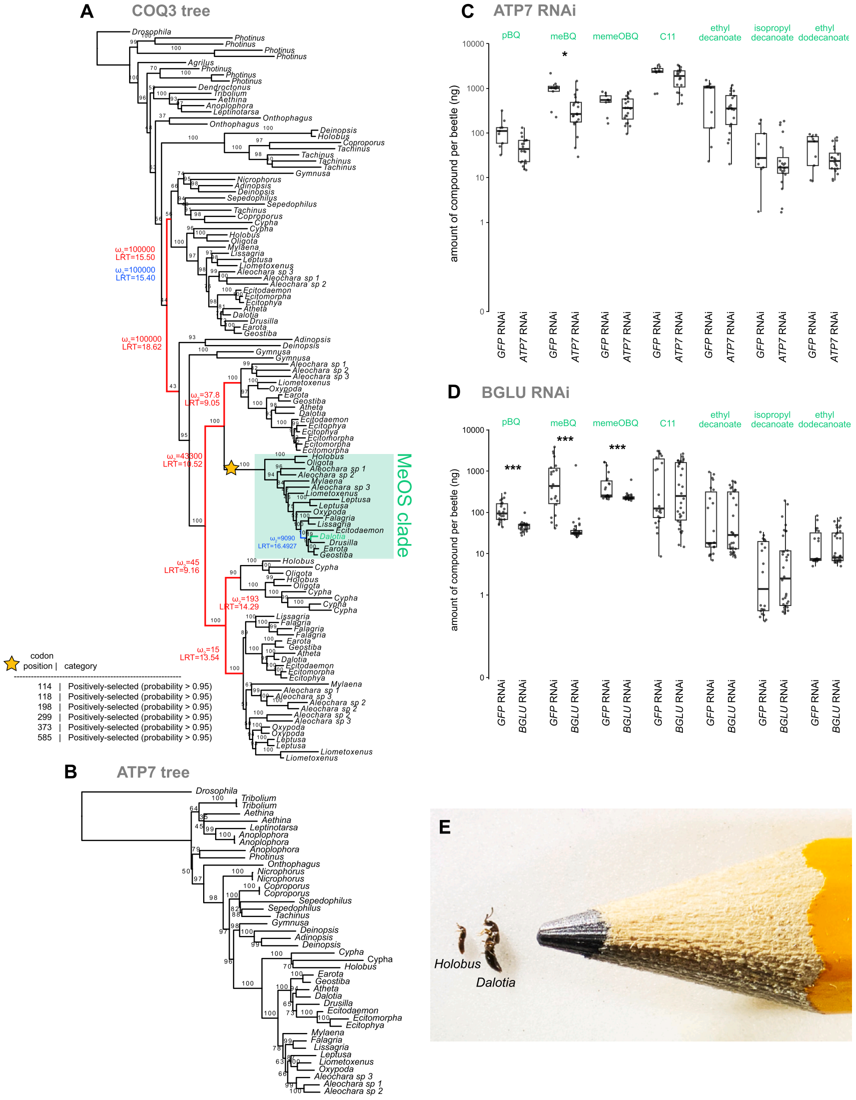


Figure S5. Evolution and function of solvent pathway enzymes, related to Figure 4

(A–C) Maximum likelihood trees of the enzymes Master Fatty Acid Synthase/MFASN (A; Q.insect+R5 model), Cytochrome P450 4G/CYP4G (B; Q.insect+R5) and Bubblegum/Bgm (C; LG + I + G4 model). Bootstrap support values are shown for each branch. *Dalotia* solvent pathway enzymes are highlighted in magenta. In B, colored branches show periods of episodic selection. aBSREL results from the all branches test are shown in red and on select branch test in blue. Associated omega (dN/dS) estimates with significant likelihood ratio test estimate (LRT) are presented for colored branches. The branch labeled for the CodeML results is indicated by a star.

(D) RNAi silencing of the very long-chain-fatty-acid-CoA ligase *bgm* in *Dalotia* selectively diminishes the levels of undecane and ethyl decanoate.

(E) HCR labeling of *MFASN* (magenta) in *Dalotia* reveals expression in solvent cells as well as fat body tissue distributed throughout the abdomen. Green: wheat germ agglutinin (WGA), which label the BQ cells.



codon position	category
114	Positively-selected (probability > 0.95)
118	Positively-selected (probability > 0.95)
198	Positively-selected (probability > 0.95)
299	Positively-selected (probability > 0.95)
373	Positively-selected (probability > 0.95)
585	Positively-selected (probability > 0.95)

Figure S6. Evolution and function of BQ pathway enzymes, related to Figure 5

(A) Maximum likelihood tree of methyoxyless/MeOs using the LG + R6 model. The MeOs clade is highlighted in the green box. Colored branches show periods of episodic selection. aBSREL results from the all branches test are shown in red and on select branch test in blue. Associated omega (dN/dS) estimates with significant likelihood ratio test estimate (LRT) are presented for colored branches. The branch labeled with a star was tested with CodeMLbranch-site model. Significant amino acid positions under selection based on Bayes Empirical Bayes analysis are presented in panel B inset table.

(B) Maximum likelihood tree of copper-transporting ATPase 1/ATP7 using Q.insect+R5 model. For each tree, bootstrap support values are shown for each branch. *Dalotia* BQ pathway enzymes are highlighted in green.

(C and D) RNAi silencing of the *ATP7* (C) and *β-glucosidase* (D) in *Dalotia* selectively diminishes the levels of benzoquinones.

(E) Photograph of *Holobus*, on the left, next to *Dalotia*, in the center, and a standard size pencil on the right.

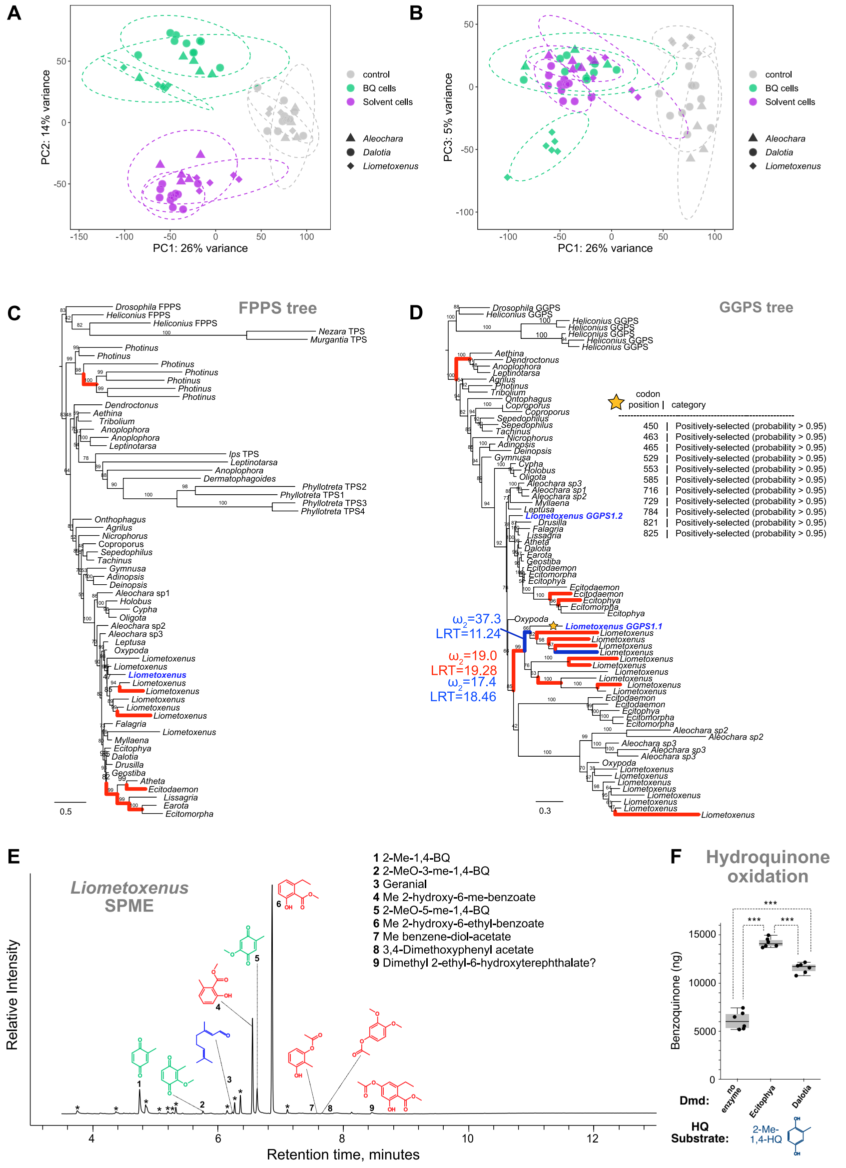


Figure S7. Evolution of myrmecophile tergal gland chemistry, related to Figure 7

(A and B) PCAs of replicate cell type specific transcriptomes from *Liometoxenus*, *Dalotia* and *Aleochara* (sp. 3) based on 8641 orthologous loci. (A) PC1 vs. PC2; (B) PC1 vs. PC3.

(C) Maximum likelihood tree of farnesyl pyrophosphate synthase/FPPS using the Q.insect+R5 model.

(D) Maximum likelihood tree of geranylgeranyl pyrophosphate synthase/GGPS using the JTT+F+I + G4 model. Bootstrap support values are shown for each branch for each tree. In both trees, colored branches show periods of episodic selection. Results of the all branches test are shown in red and on select branch test in blue with the associated omega (dN/dS) estimates and likelihood ratio test estimate (LRT) for branches leading to *Liometoxenus* genes upregulated in BQ cells (blue). The branch labeled with a star was tested with CodeMLbranch-site model. Significant amino acid positions under selection based on Bayes Empirical Bayes analysis are presented in panel B inset table.

(E) Volatilized chemicals from *Liometoxenus* glandular excretion. Headspace volatiles from a single *Liometoxenus* beetle detected via single-phase micro-extraction (SPME).

(F) Enzyme activity of Dmd from *Ecitophya*. Synthesized Dmd of *Ecitophya* can convert a 2-methyl-1,4-hydroquinone substrate (HQ) into the corresponding benzoquinone at an efficiency that exceeds that of *Dalotia* Dmd *in vitro*. Asterisks denote $p < 0.0001$ in Tukey *post-hoc* tests.