



Primary Metabolism co-Opted for Defensive Chemical Production in the Carabid Beetle, *Harpalus pensylvanicus*

Adam M. Rork¹ · Sihang Xu² · Athula Attygalle² · Tanya Renner¹

Received: 28 September 2020 / Revised: 18 December 2020 / Accepted: 2 February 2021 / Published online: 10 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021, corrected publication 2021

Abstract

Of the approximately one million described insect species, ground beetles (Coleoptera: Carabidae) have long captivated the attention of evolutionary biologists due to the diversity of defensive compounds they synthesize. Produced using defensive glands in the abdomen, ground beetle chemicals represent over 250 compounds including predator-deterrent formic acid, which has evolved as a defensive strategy at least three times across Insecta. Despite being a widespread method of defense, formic acid biosynthesis is poorly understood in insects. Previous studies have suggested that the folate cycle of one-carbon (C1) metabolism, a pathway involved in nucleotide biosynthesis, may play a key role in defensive-grade formic acid production in ants. Here, we report on the defensive gland transcriptome of the formic acid-producing ground beetle *Harpalus pensylvanicus*. The full suite of genes involved in the folate cycle of C1 metabolism are significantly differentially expressed in the defensive glands of *H. pensylvanicus* when compared to gene expression profiles in the rest of the body. We also find support for two additional pathways potentially involved in the biosynthesis of defensive-grade formic acid, the kynurenine pathway and the methionine salvage cycle. Additionally, we have found an array of differentially expressed genes in the secretory lobes involved in the biosynthesis and transport of cofactors necessary for formic acid biosynthesis, as well as genes presumably involved in the detoxification of secondary metabolites including formic acid. We also provide insight into the evolution of the predominant gene family involved in the folate cycle (MTHFD) and suggest that high expression of folate cycle genes rather than gene duplication and/or neofunctionalization may be more important for defensive-grade formic acid biosynthesis in *H. pensylvanicus*. This provides the first evidence in Coleoptera and one of a few examples in Insecta of a primary metabolic process being co-opted for defensive chemical biosynthesis. Our results shed light on potential mechanisms of formic acid biosynthesis in the defensive glands of a ground beetle and provide a foundation for further studies into the evolution of formic acid-based chemical defense strategies in insects.

Keywords Coleoptera · Formic acid · Semiochemistry · Exocrine glands · Evolution · Transcriptomics

Introduction

Arthropods are unique among animals in the diversity and novelty of chemical defensive strategies employed (Roth and Eisner 1962). Much work has been done over the years to characterize arthropod chemical defense systems both mechanistically and chemically. Functional morphologists have demonstrated how glands secrete, store, and expel chemicals, analytical chemists have characterized compounds comprising various defensive secretions, and chemical ecologists have confirmed the roles of various compounds (Arndt et al. 2016; Attygalle et al. 2020; Eisner et al. 2006; Forsyth 1972; Rork et al. 2019; Roth and Eisner 1962; Will et al. 2000). Despite this, our understanding of the biosynthetic and genetic mechanisms underlying chemical defense is absent for most taxa, as is a broader understanding of insect chemical defense evolution (Brückner and Parker 2020).

✉ Adam M. Rork
amr483@psu.edu

Sihang Xu
sxu4@stevens.edu

Athula Attygalle
athula.attygalle@stevens.edu

Tanya Renner
tur158@psu.edu

¹ Department of Entomology, The Pennsylvania State University, 501 ASI Building, University Park, PA 16802, USA

² Department of Chemistry and Chemical Biology, Stevens Institute of Technology, Hoboken, NJ 07030, USA

A particularly interesting group for studying the evolution of chemical defense and the molecular mechanisms underlying it are the ground beetles (Coleoptera: Carabidae). Members of this lineage of nearly 40,000 species are incredibly diverse in their defensive chemistries, with over 250 compounds known across hundreds of taxa (Arndt et al. 2016; Lečić et al. 2014). Among these are benzoquinones and hydrogen peroxide produced by bombardier beetles, hydrogen cyanide and benzaldehyde by tiger beetles, and methacrylic acid and salicylaldehyde by caterpillar hunters (Attygalle et al. 2020, Lečić et al. 2014, Moore and Wallbank 1968, Rork and Renner 2018). Even aquatic relatives of the Carabidae, the hydradephagans which include whirligig and predaceous diving beetles, produce various sesquiterpenes, norsesquiterpenes, aliphatic alcohols, and aldehydes (Francke and Dettner 2005; Ivarsson et al. 1996). These compounds are produced by a pair of homologous pygidial glands situated in the abdomen of all carabid taxa, which open to the terminal abdominal segment (Deuve 1993; Forsyth 1968, 1970, 1972). Chemicals are thought to be produced by a pair of secretory lobes and transported out of the structures into a long, resilin-rich collect duct (Rork et al. 2019). This collecting duct leads into a cuticular reservoir chamber surrounded by muscle, which upon agitation of the beetle will contract, expelling its components out of the body via an efferent duct. The entire gland system is resistant to chemical stress, being able to withstand 20–25% hydrogen peroxide in some species and up to 80% formic acid in others (Rossini et al. 1977; Schildknecht and Holoubek 1961).

One of the most common ground beetle defensive compounds is formic acid, a metabolite ubiquitous across life (Kanehisa and Murase 1977; Moore and Wallbank 1968; Will et al. 2000). In insects, the use of formic acid for defense has evolved at least three times independently, in Carabidae, Formicidae (Hymenoptera), and Notodontidae (Lepidoptera) (Attygalle et al. 1993, Hefetz and Blum 1978a, b, Will et al. 2000). Despite its importance as a chemical defensive agent, with dozens of behavioral assays showing it to be a potent predator deterrent, it remains unclear how insects biosynthesize defensive-grade formic acid (DGFA) (Roth and Eisner 1962; Will et al. 2010; Wilson and Regnier 1971). We define DGFA as formic acid, usually at concentrations greater than 20%, that is produced by specialized exocrine glands for defensive purposes.

A number of pathways involved in formic acid biosynthesis: the folate cycle of one-carbon metabolism, the kynurenine pathway, the methionine salvage cycle, alkane biosynthesis from fatty aldehydes, and formaldehyde oxidation (Brosnan and Brosnan 2016; Meiser et al. 2016; Sekowska et al. 2019). Although formic acid biosynthesis in the poison glands of formicine ants is linked to the folate cycle of one-carbon metabolism, it remains unknown if this is the only pathway for DGFA biosynthesis in insects (Hefetz and Blum 1978a, b). As

DGFA biosynthesis is still unstudied in most formic acid-using insect lineages, we explore the potential mechanisms in Carabidae here.

Harpalus pensylvanicus (Degeer, 1774) is an ideal model due to it being a common congener of other known formic acid-producing ground beetles (Kanehisa and Murase 1977). We aim to confirm the presence of formic acid in this species via gas chromatography-mass spectroscopy and identify candidate genes that may be involved in formic acid biosynthesis in its secretory lobes via tissue-specific transcriptomics. We hypothesize that the folate cycle of one-carbon metabolism has been independently co-opted for DGFA biosynthesis in Carabidae. Specifically, we hypothesize that genes underlying the folate cycle of C1 metabolism will be significantly differentially expressed in the secretory lobes of *H. pensylvanicus* relative to the whole body. Alternatively, folate cycle genes may have undergone duplication events leading to increased gene dosage and/or neofunctionalization, or different pathways may be involved in DGFA biosynthesis in this species altogether. Otherwise, novel DGFA synthesis genes might have evolved via gene fusion or fission, or via de novo gene generation from non-coding DNA (Tautz and Domazet-Lošo 2011).

Materials and Methods

Beetle collection *Harpalus pensylvanicus* specimens were hand collected from agricultural fields south of Ramblewood, Pennsylvania, USA (Lat: 40.7227, Long: -77.9302) on a weekly basis between 10 September 2017 and 07 October 2017. Individuals were transported to the laboratory where they were placed in terrariums with soil from the field site. Beetles in each terrarium were fed a diet of hulled millet, given moistened paper towels for water, and rocks for shelter. Paper towels were changed regularly to prevent mold growth, and dead beetles were removed from the terrariums as they were discovered.

GC-MS Analysis of Pygidial Gland Contents Two male beetles were induced to spray their gland contents into a 2-mL autosampler vial containing anhydrous dichloromethane (~200 µL) by positioning the apex of the abdomen near the opening of the vial and pinching one hind leg. Two female beetles were induced to spray into a second vial via the same method. The extracts were analyzed on a Shimadzu GC-17A fitted with a 30 m × 0.25 mm capillary column coated with ZB-WAX (0.25 µm film thickness) coupled to a QP5050 mass spectrometer. The oven temperature was held at 40 °C for 3 min and increased at 15 °C/min to a final temperature of 240 °C and held for 3 min. Each sample (1 µL) was introduced into the gas chromatograph by splitless injection.

Pygidial Gland Dissections Twelve male and eighteen female *H. pensylvanicus* specimens were induced to spray their pygidial gland contents by gently pinching a hind leg. This procedure is primarily done to empty the reservoirs, inducing the beetles to begin expressing genes necessary for biosynthesizing formic acid in order to refill them. After an hour, individual beetles were separated at the prothoracic-mesothoracic junction and stored in RNAlater™ Stabilization Solution at −80 °C until dissection.

Dissections were carried out using an Olympus SZX16 stereomicroscope. Watch glasses were sterilized with ethanol and subsequently cleaned with RNase Away Decontamination Reagent®, while forceps were flamed and decontaminated with RNase Away Decontamination Reagent® prior to dissection. Twelve male and eighteen female beetles were dissected and pooled into two tissue types for sequencing: (1) pygidial gland secretory lobe tissue and (2) whole bodies minus the secretory lobe tissue, elytra, and wings. Three biological replicates were collected in total, four male and six female beetles per replicate. Pooling was primarily done due to the small size of the glands, aiding in the recovery of ample RNA for sequencing. The elytra and wings were frozen in RNAlater™ Stabilization Solution as vouchers (Raw sequencing data is accessible via the BioProject ID: PRJNA705215).

RNA Extractions, Library Preparation, and Illumina Sequencing Total RNA was extracted from the beetle secretory lobe cells using a modified guanidinium thiocyanate-phenol-chloroform protocol (TRIzol®). All secretory lobe RNA samples were purified via a Directzol RNA MicroPrep Kit. For the whole beetle bodies without secretory lobes, a TRIzol®/chloroform/isopropanol extraction was used to extract total RNA. Six libraries were prepared using Illumina TruSeq (TR01–03) and Nextera (TR04–06) Library Preparation Kits and were sequenced on an Illumina NovaSeq 6000 (150 bp, paired-end) at a target depth aim of 50 M reads/library (Conehisa et al. 2016; Schurch et al. 2016).

Transcriptome Assembly and Quality Control Read quality was assessed for all samples using FastQC (v0.11.7) and pre-processing was done with Fastp (v0.19.4) (Chen et al. 2018). Average quality was high across all samples with very little variance, so we opted to quality trim by removing reads where over 40% of all bases were of Q20 or lower rather than using a sliding window-based approach. We also removed low complexity reads and reads not meeting the minimal inclusion length (100 bp), as well as adapter sequences, polyG, and polyA tails. Read quality was then reassessed via the report autogenerated by the Fastp program and separately via FastQC.

Given that a complete genome is currently unavailable for *H. pensylvanicus*, a transcriptome was assembled de novo via Trinity RNASEq v2.8.4 (Grabherr et al. 2011; Haas et al.

2013; Conesa et al. 2016). With the exception of read orientation being specified as “RF” and minimum contig length set to 300 bp instead of 200 bp, default settings were used for assembly. This includes in silico read depth normalization which removed reads with coverage >200. Assembly statistics were calculated as N50 and EXN50 values. The completeness of the transcriptome was assessed via benchmark single-copy ortholog (BUSCO v3.0.2) searches against a database of endopterygote sequences (endopterygota_odb9, version 13 Feb. 2016) (Simão et al. 2015).

Read Alignment and Quantification A pseudoalignment-based strategy was undertaken to quantify gene expression levels in the *H. pensylvanicus* transcriptome. Kallisto (v0.45.1) was also used to assess gene expression via its pseudoalignment-based algorithm using default settings (Bray et al. 2016). Note that without a reference genome, Trinity relies solely upon the de novo assembly process for gene classification, a process which can be prone to error, especially when using pooled, heterozygous samples.

Coding Region Prediction and Functional Annotation Due to the lack of a reference genome, we used an ab initio method of coding sequence prediction from assembled transcripts, namely the TransDecoder package (Haas et al. 2013). Default settings were used to predict likely coding sequences and translate them into corresponding amino acid sequences.

The following steps were used as part of the Trinotate annotation suite (Haas et al. 2013). To assess probable homology of the assembled transcripts and their predicted peptides, BLASTX and BLASTP (v2.6.0) were used to query the sequences against the UniProtKB database (release 2019_03) (Altschul et al. 1990). HMMScan (v3.1) was also used to annotate domains within predicted amino acid sequences using the Pfam-A HMM database (release 2018_08) (Eddy 1998; Finn et al. 2011; Punta et al. 2012; Wheeler and Eddy 2013). SignalP and TMHMM were used to predict potential coding sequences and transmembrane helices within amino acid sequences respectively (Krogh et al. 2001; Petersen et al. 2011). These data were loaded into the Trinotate boilerplate using a script provided with the software. Further fields such as KEGG pathway information, eggNOG annotations, Gene Ontology (GO) term assignments, etc., were populated based on the BLASTP and HMMScan results (Jensen et al. 2008; Kanehisa et al. 2012; The Gene Ontology Consortium 2000).

Differential Gene Expression Analyses Kallisto quantification results were used as inputs to our differential gene expression analyses. Differential gene expression analyses were conducted using the run_DE_analysis.pl and analyze_diff_expr.pl scripts provided as part of the Trinity package. Briefly, contrasts were performed between the secretory lobe tissue samples and the whole-body tissue samples. In this study, we are

primarily concerned with those genes expressed significantly higher in the secretory lobe tissues compared to the rest of the body. Voom was used to log₂-transform gene count data to log-counts per million, variances were calculated for each logCPM value, and these were transformed to quality weights for input into limma's linear model (Law et al. 2014; Ritchie et al. 2015). We chose to analyze differential expression between genes rather than transcripts due to its greater power to detect significantly DE features. Results are reported as log₂-fold change, logCPM, unadjusted *p*-values, and *p*-values adjusted via the Benjamini-Hochberg procedure for each quantification set (Benjamini and Hochberg 1995). We decided upon a critical value of $\alpha = 0.05$ for our BH-adjusted *p*-values and a log₂-fold change cutoff of 2.0.

To search for genes involved in formate biosynthesis, we used a custom AWK script to query our list of genes differentially expressed in the secretory lobes relative to the whole body. To search for differentially expressed genes involved in the folate cycle, we searched Pfam accessions specific to folate-cycle genes. These were PF00763.23 (THF DHG/CYC, catalytic domain), PF02882.19 (THF DHG/CYC, NAD(P)⁺-binding domain), PF01268.19 (Formate-THF ligase), and PF00464.19 (Serine hydroxymethyltransferase). To search for other genes involved in formate biosynthesis, we searched for eggNOG and GO Accessions containing the terms “formate”, “formic”, “forma”, and “formyl” prior to a manual search of all differentially expressed genes to identify others potentially involved in formate metabolism.

GO Enrichment Analysis By setting the flag “–examine_GO_enrichment” in the analyze_diff_expr.pl script, and by providing Trinotate-generated GO Annotations and gene lengths a GO Enrichment analysis was automatically conducted via Goseq (Young et al. 2010). Reported were enriched and depleted GO Terms for each tissue type relative to the other. The GO terms enriched in the secretory lobes relative to the whole body were further split into “Biological Process”, “Molecular Function”, and “Cellular Component” and manually analyzed. Terms descriptive of processes potentially involved in defensive chemical biosynthesis and transport were flagged for the purpose of summarizing long lists of upregulated genes. False discovery rate control was done using the Bioconductor package “qvalue” (Storey 2002).

Phylogenetic Reconstruction of Metazoan MTHFD Gene Family The MTHFD gene family comprises the core of the folate cycle of one-carbon metabolism, which we hypothesize is involved in DGFA biosynthesis in *H. pensylvanicus*. To examine the question of whether MTHFD gene family members have undergone obvious duplication events, we reconstructed the MTHFD gene phylogeny across Opisthokonta including sequences upregulated in the secretory lobes of *H. pensylvanicus*.

A custom script was used to download MTHFD gene family members from across the Opisthokont phylogeny from NCBI. An attempt was made to represent as many major phyla as possible, although the scarce genetic data for many groups meant that we could generally only obtain high-quality sequences for well-studied lineages. Sequences were aligned in MAFFT (v7.455) using (Katoh and Standley 2013). For all MTHFD1 and MTHFD1L sequences, the formyl THF synthetase domains were trimmed from the alignment leaving only the homologous THF dehydrogenase/cyclohydrolase catalytic and NAD(P)⁺-binding domains (Castresana 2000; Talavera and Castresana 2007). These domains are common to all MTHFD gene family members given their core function of converting 5,10-methylene-THF to 10-formyl-THF and vice-versa.

Phylogenetically uninformative regions were computationally identified and removed via Gblocks webserver (v0.91B) (Castresana 2000, Talavera and Castresana 2007). Phylogenetic model selection was performed with ModelFinder as part of the IQ-Tree package (v1.6.12), which recovered LG + I + G4 as the best-fit model of those tested (Kalyaanamoorthy et al. 2017; Nguyen et al. 2015). The sequence alignment file along with relevant model parameters were input to IQ-TREE with 1000 ultra-fast bootstrap (BS) replicates specified to generate branch support values. The maximum-likelihood phylogeny with branch support values was generated in FigTree (v1.4.4). The phylogeny was rooted to reflect the two distinctly paralogous clades (MTHFD1-like and MTHFD2-like) and manually annotated in Adobe Illustrator 2020 (Haque et al. 2019).

Test of Episodic Selection Acting on *Harpalus pensylvanicus*

MTHFD2 Assessment of the MTHFD2 sequence of *Harpalus pensylvanicus* and subsequent phylogenetic reconstruction revealed a number of seemingly apomorphic features relative to other holozoan MTHFD2. Assuming these are real features of the sequence and not errors, we undertook a manual analysis of an Adephaga-specific MTHFD2 protein alignment to catalog deviations from otherwise conserved residues in *H. pensylvanicus*. To determine if such deviations are the result of episodic positive selection acting on *H. pensylvanicus* MTHFD2, we also performed a branch-site test to formally test this hypothesis.

***Harpalus pensylvanicus* MTHFD2** was used as an input to a tBLASTn search of the Transcriptome Shotgun Assembly Sequence Database on the NCBI BLAST server (Altschul et al. 1990). Only hits belonging to the beetle suborder Adephaga were reported. Nucleic acid sequences were downloaded and their coding sequences were predicted using TransDecoder as described in Section 2.7. Where two proteins were predicted per transcript, the shorter of the two was discarded. Where two species in the same genus were present

in the dataset, the sequence one with the longest ORF was retained with one exception, *Noterus clavicornis*, where the longer ORF was a likely poor annotation due to low pairwise similarity to other sequences. A codon-aware alignment of the remaining sixteen sequences and *H. pensylvanicus* MTHFD2 was carried out with MACSE (v1.02) (Ranwez et al. 2011). An additional MACSE alignment was carried out with the inclusion of a human MTHFD2 sequence to aid in the identification of putatively homologous sites important for enzyme function in *H. pensylvanicus*. The resulting amino acid alignments were manually analyzed in Geneious (v11.0.5); in the human/adephagan alignment, residues highly conserved in all taxa except *H. pensylvanicus* were recorded. Gaps and the C-terminal codons were then removed from the Adephaga-specific alignment and a maximum-likelihood phylogeny of the alignment reconstructed as described in Section 2.9, with the only difference being that a codon alignment was specified (best-fit model: KOSI07 + F + R4). The codon alignment and resulting phylogeny were input into aBSREL (v2.2) to perform a branch-site test with *H. pensylvanicus* selected as the foreground (Smith et al. 2015). All other settings were left as default.

Results

GC-MS Analysis of Defensive Chemicals In both male and female *H. pensylvanicus* pygidial gland sprays, we detected formic acid, undecane, and 2-tridecanone (Fig. 1a,b, Supplementary Fig. 1a-c). Our results are in agreement with a wealth of literature on the pygidial gland contents of other *Harpalus* spp., this being the seventeenth reported formic acid-producer in the genus (Kanehisa and Kawazu 1982; Kanehisa and Kawazu 1985; Kanehisa and Murase 1977; McCullough 1966; Schildknecht 1964; Schildknecht 1961;

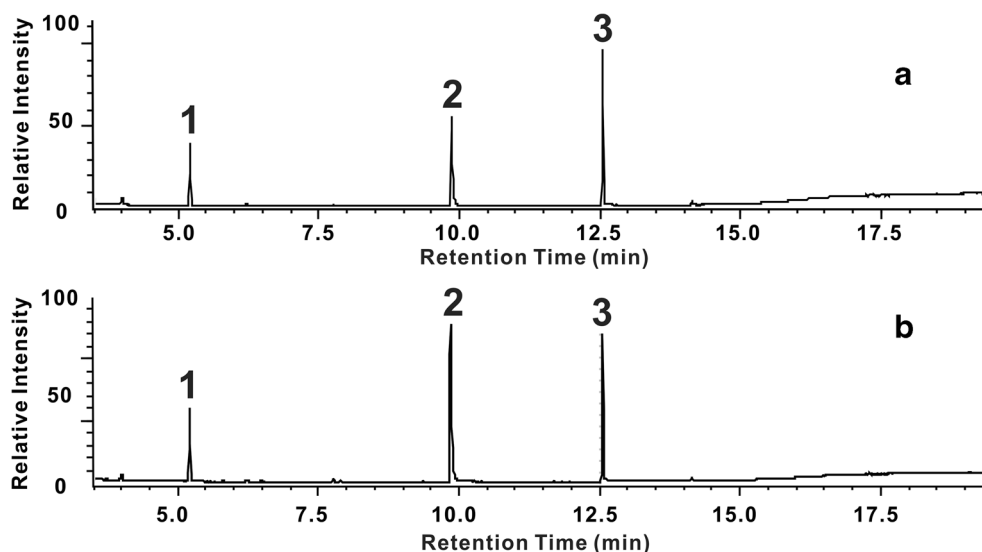
Schildknecht et al. 1968a, b; Schildknecht and Weis 1961). This is also the eighth *Harpalus* species known to produce 2-tridecanone (Kanehisa and Kawazu 1985). While we are not familiar with any literature specifically identifying undecane in the pygidial gland secretions of *Harpalus* species, it has been detected in the pygidial secretions of other carabids, often accompanying formic acid (Schildknecht 1970).

Transcriptome Sequencing Statistics We generated ~300 million total reads across six paired-end samples, totaling ~45.3 billion bp. Post-trimming, ~278 million (92.5%) of these reads remained (~41.1 billion bp) (Supplementary Table 1). The majority of trimmed reads failed the length filter, which was set based on the recommendation of Trinity-RNASeq's developers. A link to this recommendation is available in Supplementary File 1.

Transcriptome Assembly Statistics The *H. pensylvanicus* transcriptome was assembled into 160,301 transcripts (~165 million bp) which were grouped into 94,288 “genes”. The contig N50 was 1704 bp while the E90N50 was similar at 1838 bp (Supplementary Table 2). Querying the *H. pensylvanicus* transcripts against the endopterygota_odb9 database revealed a 94.1% complete transcriptome (Supplementary Table 2). Kallisto quantification results yielded pseudoalignment rates between 58.23% and 83.03% across the six samples with an average pseudoalignment rate of 70.43%.

Sequence Annotation In silico coding sequence prediction and translation analysis with Transdecoder yielded 70,782 predicted sequences from the initial transcript set. Note that while most transcripts may have exactly one predicted coding sequence, some proportion of transcripts may have more than one predicted coding sequence (Supplementary Table 3). 36.7% of transcripts received at least one BLASTX hit with an e-value <1e-5

Fig. 1 a-b Gas chromatograms generated via GC-MS analysis of the pygidial gland sprays of a male (a) and female (b) *Harpalus pensylvanicus* specimens into methylene chloride. Peak 1: Undecane; Peak 2: Formic acid; Peak 3: 2-Tridecanone. All samples run on a ZB-Wax column



against the UniProtKB database, while 70.7% of predicted peptides received a BLASTP hit. HMMscan searches annotated 66.6% of predicted peptides with at least one domain from the Pfam-A HMM. GO Annotations, linked either through BLAST or Pfam annotations, were assigned to 81.6% of predicted peptides (Supplementary Table 3).

Gene Families Involved in Formate Biosynthesis Differential expression analyses revealed 882 upregulated genes in the secretory lobes relative to the whole body. Among this set of upregulated expressed genes, we found all three genes required for the folate cycle in insects: serine hydroxymethyltransferase (SHMT), trifunctional C-1 synthetase (MTHFD1), and bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase (MTHFD2) (Figs. 2, 3). One copy of SHMT was identified in this set along with one bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase gene (MTHFD2) and three trifunctional C-1 synthetase genes (MTHFD1a-c). Alignment of MTHFD1a-c protein sequences reveals MTHFD1b and MTHFD1c to each be likely fragments of the larger MTHFD1 gene.

We also discovered two kynurenine formamidase genes, a component of the kynurenine pathway, which we denote here as KFA1 and KFA2. KFA hydrolyzes N-formyl-L-kynurenine to formate and L-kynurenine (Fig. 3, Supplementary Fig. 2). Also differentially expressed is 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase of the methionine salvage pathway, which oxidizes 1,2-dihydroxy-5-(methylsulfanyl)pent-1-en-3-one to formate and 4-(methylsulfanyl)-2-oxobutanoate (Fig. 3, Supplementary Fig. 3).

Lipophilic Molecule Biosynthesis Fatty-acyl CoA reductases (FACRs) were found to be differentially expressed in the secretory lobes of *H. pensylvanicus*, which may function in the biosynthesis of fatty alcohols from fatty-acyl CoA molecules (Fig. 3b). Fatty alcohols can be subsequently converted into n-alkanes by oxidoreductases, notably cytochrome P450s, many of which are differentially expressed in the secretory lobes of *H. pensylvanicus*. These enzymes may be involved in the biosynthesis of undecane found in *H. pensylvanicus* pygidial glands. We did not find evidence of differentially expressed aldehyde decarbonylases known to be involved in alkane biosynthesis. However, we did find evidence of differentially expressed CYP4g15, CYP4Gs broadly being known to play roles in cuticular hydrocarbon biosynthesis via oxidative decarbonylase activity (Fig. 3b) (Feyereisen 2020).

Regarding the biosynthesis of methyl ketones, we find no clear evidence of differentially expressed diketone hydrolases or decarboxylases that could convert dicarbonyl or oxo-fatty acids to methyl ketones respectively. Cytochrome P450s may also be involved in this process, as they are known to be involved in the biosynthesis of certain methyl ketones and n-alkanes in insect pheromone blends (Qiu et al. 2012).

Transport of Molecules Associated with Formate Biosynthesis

There are multiple genes differentially expressed in the secretory lobes that encode transport proteins associated with processes putatively involved in DGFA biosynthesis. A well-represented set of such transport proteins is the major facilitator superfamily (MFS) (Quistgaard et al. 2016). Of these, we found numerous organic cation and anion transporters,

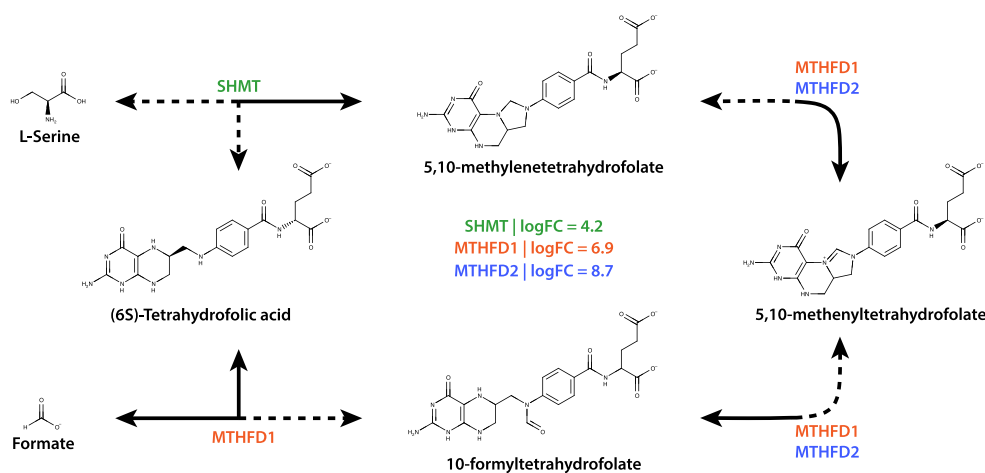


Fig. 2 Simplified representation of a portion of the folate cycle of one-carbon metabolism in insects. Solid arrows represent an oxidative direction of pathway flow where the primary output is formate and tetrahydrofolate. Dashed arrows represent a reductive direction of pathway flow where L-serine is a primary output along with tetrahydrofolate. Genes and their respective log fold change values in *Harpalus pensylvanicus* secretory lobes vs. whole body tissue are shown in the middle of the diagram. SHMT catalyzes the reversible conversion of L-Serine or L-Glycine and tetrahydrofolate to 5,10-methylenetetrahydrofolate.

MTHFD2 catalyzes the reversible dehydrogenation of 5,10-methylenetetrahydrofolate into 5,10-methenyltetrahydrofolate and the reversible hydrolysis of 5,10-methenyltetrahydrofolate to 10-formyltetrahydrofolate. MTHFD1 is capable of catalyzing the same reactions as MTHFD2 as well as the reversible generation of formate and tetrahydrofolate from 10-formyltetrahydrofolate. Not shown for simplicity are multiple cofactors and shunts from the pathways that are involved in the biosynthesis of purines, thymidylate, the regeneration of methionine, etc.

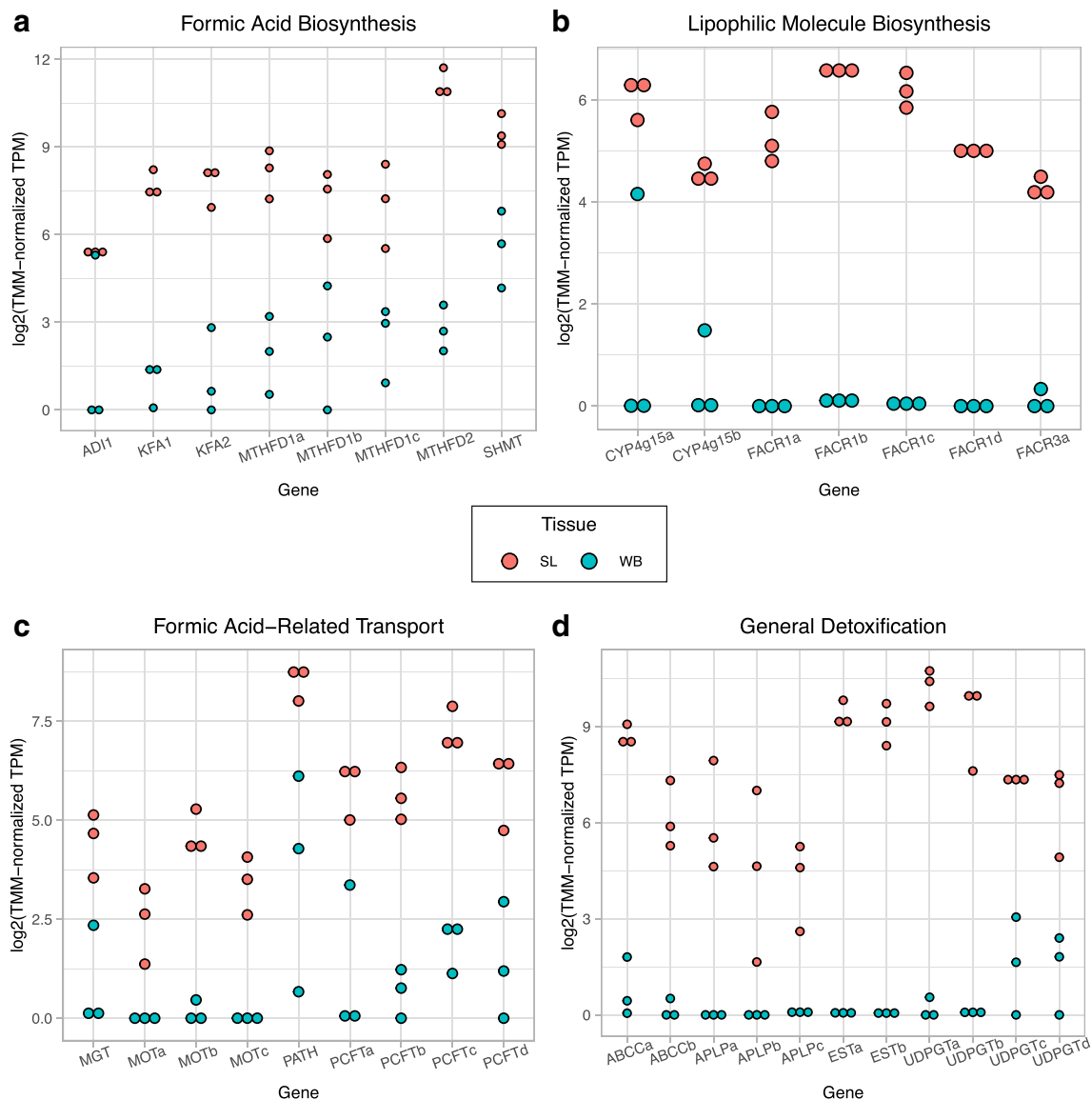


Fig. 3 a–d Dot plots demonstrating the expression levels of genes discussed as part of this study as log₂-transformed TMM-normalized TPM values. Each TPM value was increased by one, such that replicates with zero TPM could be normalized. Gene names follow the naming convention used throughout the manuscript with multiple copies/gene family members being denoted by the lower-case subscripts “a–z”. Red

dots represent secretory lobe bioreplicates while blue dots represent glandless body bioreplicates. For conciseness, not all upregulated genes of a particular family or functional group may be represented. A full list of genes, their expression values, etc., are available in Supplementary Table 4

monocarboxylate transporters (MOTs), and proton-coupled folate transporters (PCFTs) (Fig. 3c). Despite the name, organic cation transporters are also capable of transporting anions and zwitterions, thus making them as well as the monocarboxylate and anion transporters capable of transporting formate out of the cell. The folate transporters may be allowing for an increased uptake of folate into the cell to serve as an input to the folate cycle once fully reduced. There are also some monocarboxylate transporters not within the major facilitator superfamily.

We also identified multiple amino acid transport genes, varying in their substrate specificities. Among these are

mitochondrial glycine transporters (MGTs) and proton-coupled amino acid transporters (PATH) which may serve as means by which folate cycle precursors are transported into the cell and/or further into the mitochondria (Fig. 3c).

General Detoxification and Transport We found multiple gene families that may be involved in the breakdown and/or transport of other cytotoxic metabolites out of the cell. These include uridine 5′-diphospho-glucuronosyltransferase (UDPGT) genes differentially expressed in the secretory lobes, which could be responsible for glucuronidation of lipophilic metabolites (Fig. 3d) (Mulder 1992). These

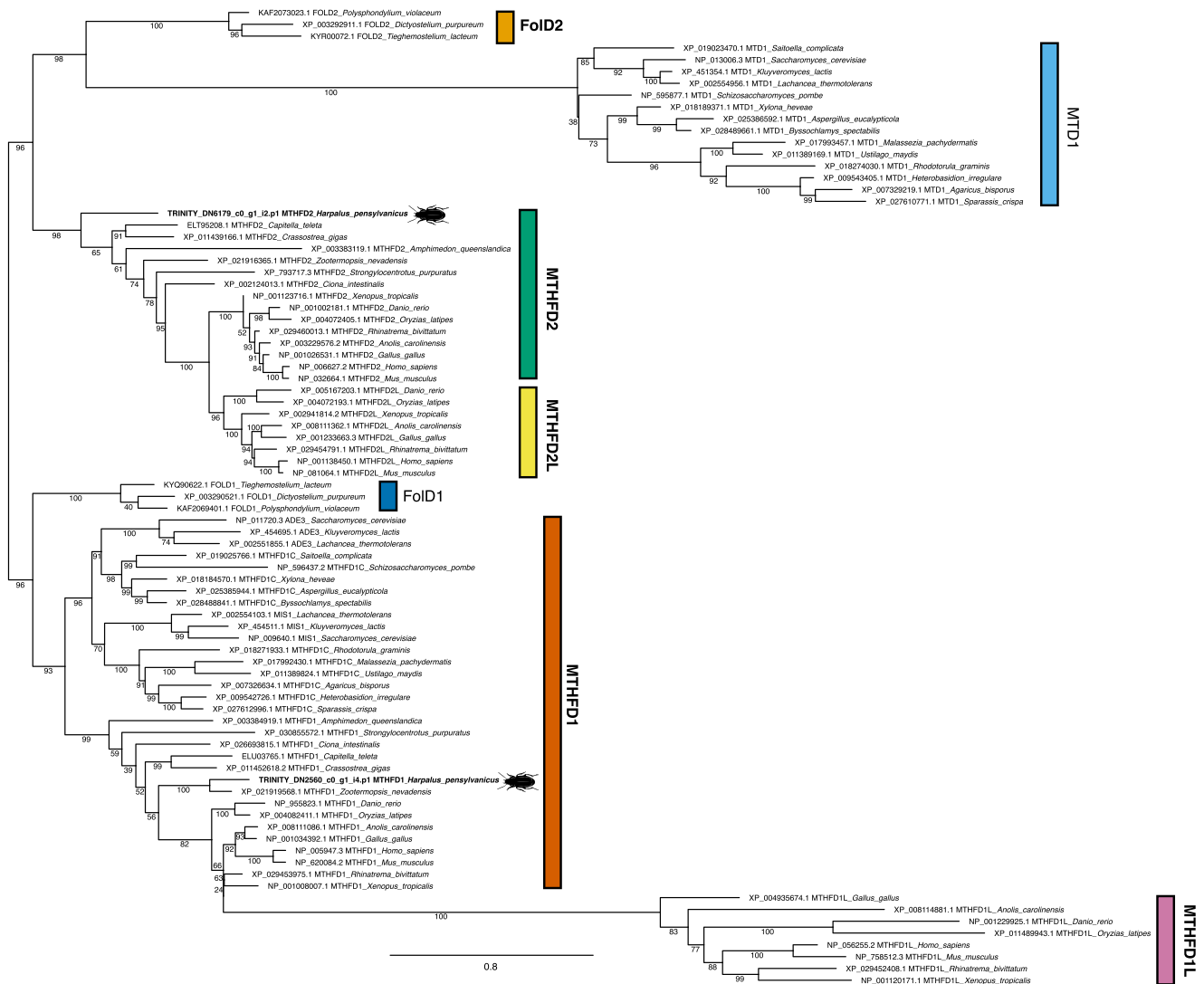


Fig. 4 Maximum-likelihood phylogeny of the MTHFD gene family with branch support values shown. Phylogeny was generated from a protein alignment of the THF/DHG/CYC catalytic and NAD(P)⁺-binding domains of MTHFD gene family members across Amorphea using IQTREE v1.6.12. Branch support values were generated via ultrafast bootstrap approximation (1000 replicates). Colored bars to the right of major clades are shown for ease of navigation. The Amoebozoa-specific

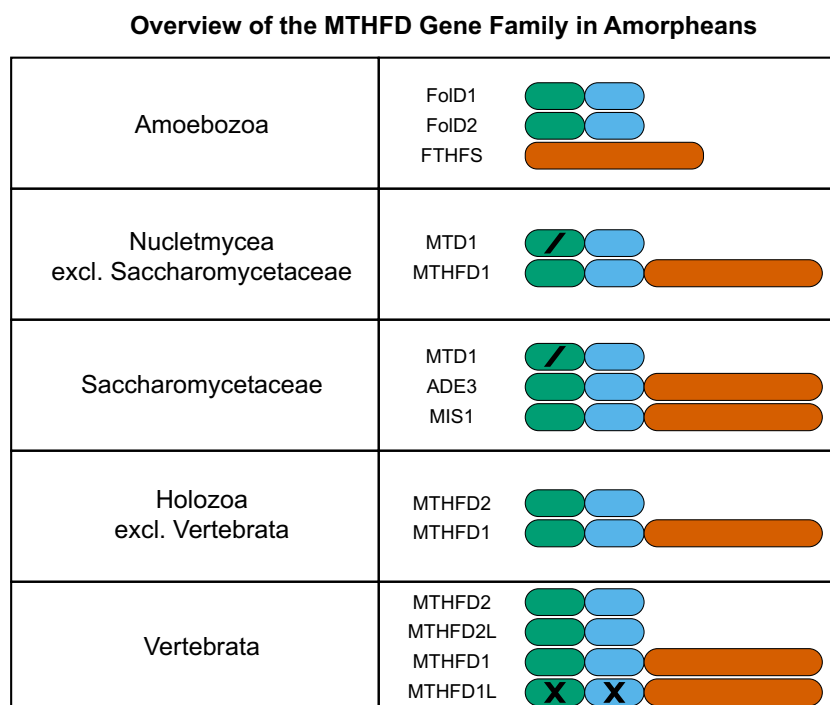
Fold1 and Fold2 clades are represented by the dark blue and light orange bars respectively. The Nuclemycea-specific MTD1 clade is represented by the light blue bar. The Holozoa-specific MTHFD2 grade is represented by the green bar and the Vertebrata-specific MTHFD2L grade by the yellow bar. The Opisthokonta-specific MTHFD1 grade is represented by the dark orange bar while the Vertebrata-specific MTHFD1L clade is represented by the pink bar

glucuronidated metabolites (as well as their non-glucuronidated forms) may be shuttled across the membrane by membrane-bound transport proteins, such as the ATP-binding cassette (ABC) and other transmembrane transport proteins differentially expressed in the secretory lobes (Fig. 3d) (Locher 2009). Transmembrane transport proteins such as those belonging to the ATP-binding cassette family, major facilitator superfamily, or solute carrier families are important in a variety of organisms for detoxification and general transport processes (Locher 2009; Quistgaard et al. 2016). We also found apolipophorins (APLPs), which may specifically be transporting lipophilic molecules into and/or out of the

cells, such as those found in the pygidial defensive fluid. In addition, there are multiple differentially expressed carboxylesterases (EST) that may be hydrolyzing ester, thioester, or amide-containing metabolites generated by the secretory lobes, as well as cytochrome P450s (CYP4) that may be acting upon a variety of substrates, oxidizing or reducing them.

GO Enrichment Our GOSeq analysis revealed 128 GO Terms enriched in the secretory lobes, 71 “Biological Processes”, 21 “Cellular Components”, and 36 “Molecular Functions”. Particularly interesting amongst the enriched biological

Fig. 5 Overview of basic MTHFD gene family architecture in amorphean taxa. The THF/DHG/CYC catalytic and NAD(P)⁺-binding domains are shown in green and blue respectively. The formyltetrahydrofolate synthetase domain is depicted in orange. A single black slash through a domain indicates altered function whereas a black X indicates lack of functionality. Gene fusion, duplication, subfunctionalization, neofunctionalization, and occasionally gene death are prominent features of MTHFD gene family evolution



processes are a series of terms related to folate metabolism. Specifically, these include “tetrahydrofolate interconversion”, “tetrahydrofolate metabolic process”, “folic acid-containing compound metabolic process”, “pteridine-containing compound metabolic process”, and “one-carbon metabolic process”. There are also enriched terms for “monocarboxylic acid transport”, “organic acid transport”, and “carboxylic acid transport”. Together, these enriched terms allude to the notion that *H. pensylvanicus* secretory lobes are sites of one-carbon metabolic activity that function in the transport of (mono) carboxylic acids, such as formic acid. Although broader, enriched terms such as “lipid biosynthetic process” and “lipid transport” may allude to the biosynthesis and secretion of compounds such as 2-tridecanone and undecane.

All aforementioned biological processes have analogous molecular function. Pertaining to folate metabolism and formate biosynthesis specifically are the enriched terms “methylenetetrahydrofolate cyclohydrolase activity”, “methylenetetrahydrofolate dehydrogenase (NADP⁺) activity”, and “formate-tetrahydrofolate ligase activity”, which describe every catalytic function leading from 5,10-methylenetetrahydrofolate to the biosynthesis of formic acid and the regeneration of tetrahydrofolate. Analogous transport functions are also enriched, such as “organic anion transmembrane transporter activity”, “organic acid transmembrane transporter activity”, and “carboxylic acid transmembrane transporter activity”, and “monocarboxylic acid transmembrane transporter activity” all of which describe a means via which formate (an anion) may be transported out of the secretory lobes. Indeed, it’s certainly the case that export to or import from the extracellular region

is important in *H. pensylvanicus* secretory lobes, as numerous such terms are enriched in the cellular component category, including “extracellular region”, “extracellular space”, “cell surface”, and “extracellular matrix”. A full list of enriched GO terms can be found in Supplementary Table 5.

The secretory lobes having numerous enriched GO terms for the one-carbon metabolic process suggests that the process is important for tissue function. In the ecological context of the secretory lobes, we hypothesize as above this pathway to be involved in defensive-grade formic acid biosynthesis. The abundance of enriched GO terms for organic acid and organic anion transport suggests that such compounds (e.g. formic acid) are likely being secreted to the extracellular space (e.g. the collecting duct). This tissue also has substantial lipid biosynthesis and transport functions, likely related both to those compounds present in gland secretions and those used for other cellular metabolic processes.

Phylogenetic Reconstruction of Amorphean MTHFD Gene Family To determine the possible role that gene duplication may have played in the evolution of formic acid biosynthesis, we reconstructed the phylogeny of amorphean MTHFD gene family members (Fig. 4, Supplementary Table 6). From our phylogenetic analysis of the homologous THF/DHG/CYC catalytic and NAD(P)⁺-binding domains of amoebozoan, animal, and fungal MTHFD gene family members, it is evident that an ancient duplication event in the MTHFD gene family occurred prior to the origins of Opisthokonta (Fig. 4). The topology of our phylogeny suggests that an ancestral copy of the gene *FoID*, which contains both THF/DHG/CYC

catalytic and NAD(P)⁺-binding domains, duplicated in a common ancestor of Amorphea (Amoebozoa + Holozoa + Nucleotmycea) (Fig. 4). In the Amoebozoa, both copies have been retained, each with apparent THF/DHG/CYC catalytic and NAD(P)⁺-binding activities, although this has never been functionally validated (Fig. 4). We recover one of two *Fold* proteins, here denoted as *Fold1*, to be sister to the THF/DHG/CYC catalytic and NAD(P)⁺-binding domains of MTHFD1 and its paralogues (96% BS). The other *Fold* protein, *Fold2*, forms a clade with MTHFD2 and its paralogues (96% BS), although its exact placement within this clade is contentious likely due to long branch attraction (Figs. 4, 5) (Küke et al. 2012, Philippe et al. 2005).

MTHFD1 and its Paralogues In a common ancestor of Opisthokonta, a *Fold1*-like paralogue appears to have fused with an ancestral single-copy of formyltetrahydrofolate synthetase (*FTHFS*) (Fig. 5). The gene product resulting from this fusion event is what is referred to in animal systems as MTHFD1. Indeed, the THF/DHG/CYC catalytic and NAD(P)⁺-binding domains of *MTHFD1* are homologous to those of an ancestral *Fold* whereas the formyltetrahydrofolate synthetase domain of *MTHFD1* is homologous to that of an ancestral *FTHFS*. In a common ancestor of Vertebrata, *MTHFD1* has undergone its own duplication event. One copy, retaining the abbreviation *MTHFD1*, has ancestral THF/DHG/CYC catalytic, NAD(P)⁺-binding, and formyltetrahydrofolate ligase activities. The other, *MTHFD1L*, has lost THF/DHG/CYC catalytic and NAD(P)⁺-binding activities through the accumulation of deleterious mutations and only functions as a formyltetrahydrofolate synthetase. An independent *MTHFD1* duplication event also occurred in some Saccharomycotina, likely specific to the Saccharomycetaceae, both paralogues having retained their trifunctional nature. In the fungal literature, these genes are referred to as *ADE3* and *MIS1* and differ primarily in their cellular localization (*ADE3* being cytoplasmic, *MIS1* mitochondrial). These clades are both strongly supported (100% BS each). The larger Opisthokont (MTHFD1 + MTHFD1L) clade has strong (93% BS) support as does the MTHFD1L subclade (100% BS). Note that although we refer to most nucleotmycean copies of MTHFD1 as MTHFD1C, these are orthologous to the holozoan MTHFD1.

MTHFD2 and its Paralogues A *Fold2*-like paralogue appears to have accumulated deleterious mutations in nucleotmyceans, rendering it incapable of performing cyclohydrolase activities. Thus, it is referred to in the literature as monofunctional methylenetetrahydrofolate dehydrogenase (*MTD1*) (West et al. 1993; Wagner et al. 2005). In Holozoa, this paralogue is generally referred to as *MTHFD2* and retains both ancestral THF/DHG/CYC catalytic and NAD(P)⁺-binding activities. In Vertebrata, interestingly, this gene has also undergone its own gene duplication event, giving rise to the paralogue

MTHFD2L. The two paralogues both retain THF/DHG/CYC catalytic and NAD(P)⁺-binding activities and differ primarily in their spatiotemporal expression patterns and relative preferences for NAD⁺ and NADP⁺. The (MTHFD2 + MTHFD2L) clade has strong support (100% BS), as does the MTHFD2L subclade (96% BS). The MTD1 clade itself has strong support (100% BS). Its placement as sister to *Fold2*, despite seemingly strong support (98% BS) is likely a consequence of long branch attraction rather than recent common ancestry (Fig. 4) (Kück et al. 2012, Philippe et al. 2005). We propose that like the MTHFD1 clade, *Fold2* may be sister to an opisthokontan MTD1/MTHFD2/MTHFD2L clade.

***Harpalus pensylvanicus* MTHFD Gene Family Members**

Within *H. pensylvanicus*, we find no substantial evidence for a lineage-specific expansion and divergence of MTHFD gene family members. That is, *H. pensylvanicus* has the typical suite of MTHFD gene family members, one probable copy of *MTHFD1* and one probable copy of *MTHFD2*. However, an expansion or lack thereof cannot be confirmed with a transcriptome alone. For example, members of the *MTHFD1a-c* clade may be transcripts originating from recently duplicated genes that still share substantial sequence similarity. Alternatively, they could be alternate transcripts from a single locus that Trinity classified as different genes. Thus, a genome is necessary to confidently examine the possibility of there being multiple MTHFD1, etc., loci.

We propose this latter scenario, alternative transcripts being classified as different genes, to be more likely for MTHFD1. Multiple sequence alignment of the MTHFD1a-c peptides and coding regions show that MTHFD1b and MTHFD1c are fragments of a larger MTHFD1 sequence. At the protein level, the homologous regions of MTHFD1a and MTHFD1b are identical, as are the homologous regions of MTHFD1a and MTHFD1c. At the coding sequence level, MTHFD1a and MTHFD1b share 99.8% pairwise similarity (2428/2434) while MTHFD1a and MTHFD1c share 100% pairwise similarity. Allelic variation present in the raw reads likely caused MTHFD1b and MTHFD1c to assemble separately from MTHFD1a, resulting in their classification as separate Trinity genes.

The full copy of *H. pensylvanicus* MTHFD1 is recovered with robust support (100% BS) as sister to its orthologue in the termite species *Zootermopsis nevadensis*, the sequences having ~71% pairwise similarity. Given that *H. pensylvanicus* and *Z. nevadensis* are the only two arthropod species in our phylogeny, this is not unexpected. More unusual is the placement of the *H. pensylvanicus* MTHFD2 as sister to the remainder of the [MTHFD2 + MTHFD2L] clade, although with weak support (65% BS). Indeed, *H. pensylvanicus* and *Z. nevadensis* MTHFD2 orthologues share ~53% pairwise similarity. While it is not surprising that gene-level

phylogenies do not necessarily correlate with species-level phylogenies, it does raise the possibility that MTHFD2 in *H. pensylvanicus* has acquired non-synonymous mutations that makes its exact phylogenetic placement within the MTHFD2 grade uncertain.

***H. pensylvanicus* MTHFD2 Is Not under Positive Selection and Has Few Novel Features**

The results of our branch-site test in aBSREL found no evidence that the branch leading to *H. pensylvanicus* MTHFD2 is under episodic diversifying positive selection ($p = 0.35688$) (Supplementary Data 1) (Smith et al. 2015). However, there are some notable deviations from otherwise evolutionarily conserved amino acid sites in the MTHFD2 enzyme of *H. pensylvanicus*. Using the coordinates in the human MTHFD2 enzyme (NP_006627.2) as a reference, we found five sites in *H. pensylvanicus* MTHFD2 that have had an amino acid substitution relative to human and all other adephagan sequences (Supplementary Data 1). These include Leu114→Met112, Leu192→Val190, Thr244→Cys242, Val250→Ile248, and Thr324→Cys320 (where *H. sapiens*→*H. pensylvanicus*). We also noted convergence between human and *H. pensylvanicus* on Val279→Val277, a site that in all other adephagans encodes for isoleucine (Supplementary Data 1). Although certainly possible, we do not suggest that these novel features have a significant bearing on enzyme function, as none have been implicated in catalytic processes to our knowledge.

Discussion

Arthropod exocrine gland systems have evolved to serve defensive purposes in a variety of taxa. The Carabidae represent a unique case where multiple distinct chemical defense strategies have evolved across the family, all utilizing a homologous gland system to produce and secrete these compounds (Forsyth 1972; Moore and Wallbank 1968; Rork and Renner 2018). While much work has been done to characterize these compounds over the past century, we still know little about the biosynthesis and genetic basis for many of these compounds. Here, we have provided evidence for the production of DGFA in the pygidial gland system of *H. pensylvanicus* and discuss possible scenarios for its biosynthesis based on our tissue-specific transcriptomic analyses.

Similar to other species of *Harpalus*, we confirmed that *H. pensylvanicus* produces formic acid as a primary defensive compound in its pygidial glands (Kanehisa and Kawazu 1985). Formic acid is a common pygidial gland allomone in the Carabidae, being found in dozens of genera in three different subfamilies (Harpalinae, Trechinae, and Psydriinae) (Attygalle & Will, unpublished data, Kanehisa and Kawazu 1982, 1985, Kanehisa and Murase 1977, Moore and Wallbank 1968, Rossini et al. 1977). *H. pensylvanicus* also produces 2-

tridecanone, as well as undecane, an alkane found in the sprays of other formic acid-producing Carabidae (Kanehisa and Kawazu 1985; Schildknecht 1970). These compounds may play a role in enhancing the irritating effects of formic acid by acting as an agent to penetrate the hydrophobic cuticles, as is thought to be the case in other chemically defended arthropods (Eisner et al. 1961; Peschke and Eisner 1987). Alternatively, the polar organic constituents could serve as solvents to deliver formic acid. There are at least four genera in the subfamilies Harpalinae and Psydriinae that have a similar pygidial gland composition to *H. pensylvanicus* (at least formic acid + undecane + 2-tridecanone), suggesting multiple instances of convergent or parallel evolution upon a specific chemical blend (Moore 1979).

Defensive-Grade Formic Acid Biosynthesis May Involve Three Distinct Pathways

The results of our tissue-specific RNA-Seq experiments and subsequent differential expression analyses suggest one to three pathways may be contributing to DGFA production in the pygidial glands of *H. pensylvanicus*. All of the genes involved in the folate cycle of one-carbon metabolism in invertebrates are over-expressed in the secretory lobes of *H. pensylvanicus* (Brosnan and Brosnan 2016; Meiser et al. 2016). In *Drosophila melanogaster*, *Caenorhabditis elegans*, and others, these genes are serine hydroxymethyltransferase (SHMT), bifunctional 5,10-methylenetetrahydrofolate dehydrogenase/cyclohydrolase (MTHFD2), and trifunctional C-1 synthetase (MTHFD1). SHMT catalyzes the simultaneous conversion of L-serine to L-glycine and tetrahydrofolate (THF) to 5,10-methylene-THF. 5,10-methylene-THF is subsequently converted to 5,10-methenyl-THF, which is hydrolyzed to 10-formyl-THF by the dehydrogenase/cyclohydrolase domains of MTHFD1 or MTHFD2 (Fig. 2). The ligase domain of MTHFD1 then cleaves the formyl group from 10-formyl-THF, regenerating tetrahydrofolate and a formate anion. This pathway is also reversible, capable of generating L-serine and THF from formate and THF. We also identified two differentially expressed copies of 5-formyltetrahydrofolate cyclo-ligase, which can generate 5,10-methenyl-THF from 5-formyl-THF, which is itself produced from 5,10-methenyl-THF via SHMT (Fig. 2). The role of MTHFS in this cycle as it pertains to DGFA biosynthesis is unknown, but it may play a role in regulating the rate of reaction by consuming and regenerating 5,10-methenyl-THF. It is intriguing that none of the genes involved in the usual biosynthetic shunts from this pathway are differentially expressed. Across life, a major function of the folate cycle is to generate one-carbon units for the biosynthesis of deoxyribonucleoside, with 10-formyl-THF being a substrate for GAR formyltransferase and AICAR formyltransferase (purines) and 5,10-methylene-THF being a substrate for thymidylate synthase (thymidine) (Carreras and Santi 1995; Fox and Stover 2008; Hartman and Buchanan 1959). We do

not observe these or other genes involved in these pathways as differentially expressed. This indicates that a majority of the one-carbon units generated may be more likely to contribute to the biosynthesis of formate anions than deoxyribonucleosides.

The second step in the kynurenine pathway may also be generating DGFA from N-formylkynurenine (Badawy 2017). Only the kynurenine formamidase (*KFA*) genes were found to be differentially expressed in the secretory lobes, which hydrolyze the amide bond of N-formylkynurenine, generating a formate anion and kynurenine (Han et al. 2012; Mehler and Knox 1950). However, given that tryptophan-2,3-dioxygenase (TDO), the preceding enzyme in the pathway, is not differentially expressed, we suggest that *KFA* may only play a minor role in formate production. The kynurenine pathway is rate-limited by TDO, making it unlikely that enough N-formylkynurenine could be generated to produce a significant amount of formate (Kanai et al. 2009). That, and no genes coding for downstream enzymes in the pathway were differentially expressed, suggesting that there could be a build-up of kynurenine after it is generated from kynurenine formamidase. Also, given that tryptophan is an essential amino acid, we think it is unlikely that *H. pensylvanicus* would be using large quantities of a valuable metabolic resource for DGFA production.

The methionine salvage cycle, like the kynurenine pathway, may only have a minor role in DGFA biosynthesis if any for many of the same reasons. Only 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase, the enzyme that generates formate and 4-(methylsulfanyl)-2-oxobutanoate in the cycle, was found to be differentially expressed. Rate-limiting steps would still likely be incapable of producing enough product for 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase to convert to substantial quantities of DGFA without increased expression (Savarse et al. 1985). As with tryptophan, given the status of methionine as an essential amino acid, we do not consider it likely that *H. pensylvanicus* is using large quantities of this compound for formate biosynthesis even though methionine can be regenerated from homocysteine after it is used in the salvage cycle (Froese et al. 2019; Jaffe and Chrin 1979; Parkhitko et al. 2019).

We suggest that the kynurenine pathway and methionine salvage cycle may be playing a minor role in DGFA biosynthesis, where the folate cycle plays a more central role. The differential expression of all genes central to the folate cycle and complementary substrate transporters, the likely greater availability of L-serine and L-glycine compared to L-tryptophan and L-methionine, and the evidence that we have from Formicidae DGFA biosynthesis make the folate cycle of one-carbon metabolism a promising candidate (Hefetz and Blum 1978a, 1978b). This is also supported by our GO enrichment analyses, which show processes and functions directly tied to one-carbon metabolism through the folate cycle

to be enriched in the secretory lobe tissue compared to gland-less body tissue.

Secretory Lobes Are a Hub of Secondary Metabolite Biosynthesis & Detoxification Our differential gene expression analyses show multiple differentially expressed genes in the secretory lobes that are known to be involved in the chemical modification and transport of various metabolites in other taxa. Among these are the UDPGTs, one copy of which is the most highly expressed gene in the secretory lobes. UDPGTs chemically modify what are often large, lipophilic molecules via the addition of a glucuronic acid molecule. These glucuronidated molecules (aka glucuronides) tend to be more hydrophilic than their non-glucuronidated counterparts, allowing them to be expelled from cells via transport proteins more efficiently (Mulder 1992). What substrates these enzymes may be working upon is currently unknown. Additionally, we found carboxylesterases and cytochrome P450s that could play a role in chemically modifying cytotoxic byproducts of a metabolically active tissue, namely the secretory lobes (Coppin et al. 2012; Jackson et al. 2013; Schuler and Berenbaum 2013; Scott and Wen 2001). Some of this chemical modification of metabolites may lead to the biosynthesis of 2-tridecanone and undecane, although this remains uncertain. The multiple ABC, MFS, and other transport proteins found to be differentially expressed in the secretory lobes suggest a general necessity to transport small, polar molecules such as formate out of the cell while transporting molecules such as folate and amino acids into the cell (Locher 2009; Quistgaard et al. 2016). This is reflected in numerous transport-related GO terms enriched in the secretory lobes, including the term “transport” itself, lending credence to the notion that the folate cycle may be a fundamentally important process in the secretory lobes due to its role in formate biosynthesis. Indeed, the presence of genes involved in small molecule transport, including in the context of chemical defense, has been noted in other gland-specific transcriptomes (Brückner and Parker 2020). Future isotope tracing studies are needed to provide evidence for involvement of the folate cycle in DGFA biosynthesis (Attygalle et al. 2006; Attygalle et al. 2020). Although useful in linking genes to phenotypes, gene knockdown/knockout studies may be too deleterious to be useful in this context, given the fundamental role of these genes in primary metabolic processes (Di Pietro et al. 2002; MacFarlane et al. 2009).

The Evolutionary Complexity of the MTHFD Gene Family The MTHFD gene family is rather mechanistically complex, as is its evolutionary history, and dates back to the origin of life. We suggest that the most recent common ancestor of Opisthokonta had what was functionally a single copy of MTHFD1 and a single copy of MTHFD2.

We recovered two distinct clades following phylogenetic analyses of the homologous THF/DHG/CYC catalytic and NAD(P)⁺-binding domains of animal and fungal MTHFD gene family members: one specific to MTHFD2, the paralogous MTHFD2L, and MTD1 as well as one specific to MTHFD1 and the paralogous MTHFD1L (Figs. 4, 5). The addition of *H. pensylvanicus* MTHFD gene family members into the phylogeny does not suggest any obvious gene duplication events within this species, although confirmation of this requires thorough investigation with a high-quality, well-annotated genome. This is important since transcriptomes may not accurately represent the full gene space of an organism and do not necessarily contain enough information to confidently distinguish between isoforms and recently-duplicated genes. These findings, in the context of our gene expression analyses, implies that regulation of expression rather than gene duplication and/or possible neofunctionalization events are more important for folate cycle upregulation in the secretory lobes of this ground beetle species. This is further supported by a lack of evidence that MTHFD2 of *H. pensylvanicus* is under episodic diversifying positive selection. This is not necessarily surprising, given the central role of MTHFD2 in primary metabolism. While there are certainly some unique features of *H. pensylvanicus* MTHFD2 compared to relatively closely related adephagans, due to the lack of knowledge regarding the structure and function of insect MTHFD2, few if any conclusions can be reliably drawn as to their functional significance, especially as they may pertain to DGFA biosynthesis (Gustafsson et al. 2017).

While the mechanisms are currently unknown in *H. pensylvanicus*, we suggest either selection on duplicated regulatory elements or tissue-specific transcription factors involved in *MTHFD1* and *MTHFD2* upregulation (Arsovski et al. 2015; Li and White 2003; Long et al. 2016; Smith et al. 2013; Sonawane et al. 2017; Steuernagel et al. 2019; Wray 2007). Unfortunately, we know little about the putative cis-regulatory elements or their transcription factors for *MTHFD1* or *MTHFD2* in insects. Most of what is known is from human systems, which may or may not be applicable to insect systems (Ben-Sahra et al. 2016; Carroll et al. 2009). Given that MTHFD gene family members are constitutively expressed in all body tissues of most animals, their uniquely high expression the secretory lobes of *H. pensylvanicus* provides for a unique system in which the regulation of MTHFD gene family members and the folate cycle may be studied.

Conclusions Our results suggest that DGFA biosynthesis in the pygidial glands of *H. pensylvanicus* may be primarily linked to one to three primary metabolic pathways: the folate cycle of one-carbon metabolism, the kynurenine pathway, and the methionine salvage cycle. We suggest that increased expression of key genes in each pathway may be contributing to the enhanced ability of *H. pensylvanicus* to generate relatively

large, concentrated quantities of formic acid that they store and secrete. Of these three pathways, the folate cycle is particularly promising as a candidate pathway, given that all genes in this pathway are differentially expressed and it was found to be the means of DGFA biosynthesis in formicine ants (Hefetz and Blum 1978a, 1978b). These results aid in our understanding of the evolution of insect chemical defense systems and provide evidence for a parallel evolutionary mechanism of DGFA biosynthesis in two distinct insect lineages. Future studies will focus on understanding the role of each pathway in DGFA biosynthesis via radiolabeled precursor injections and the expansion of this work to other formic acid-defended ground beetle taxa.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10886-021-01253-2>.

Acknowledgements We thank Yuka Imamura and Kimberly Moran for their assistance with RNA extraction, library preparation, and RNA-Seq. We would also like to thank Wendy Moore, Reilly McManus, Kipling Will, Sean Perez, and Aman Gill for thoughtful discussion on Carabidae biology and chemical defense mechanisms. Additionally, we thank Wendy Moore, Kipling Will, Kylie Bocklund, Chloe P. Drummond, Katie Nolan, Kadeem Gilbert, Arthi Bala, Dana Roberts, Jonah Ulmer, and István Mikó for discussion and critique of this project, its results, and interpretation. This work was supported by the National Science Foundation [DEB 1762760, 1556898].

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Arndt E, Beutel RG, Will K (2016) Carabidae Latreille, 1802. In: Beutel RG, Leschen RAB (eds) handbook of zoology. Vol. IV Arthropoda: Insecta, part 38. Coleoptera, beetles. Vol 1: morphology and systematics (Archostemata, Adephaga, Myxophaga, Polyphaga partim). Walter de Gruyter, Berlin, pp 119–144
- Arsovski AA, Pradinuk J, Guo XQ, Wang S, Adams KL (2015) Evolution of Cis-regulatory elements and regulatory networks in duplicated genes of Arabidopsis. *Plant Physiol* 169(4):2982–2991
- Attygalle AB, Smedley SR, Meinwald J, Eisner T (1993) Defensive secretion of two notodontid caterpillars (*Schizura unicornis*, *S. badia*). *J. Chem Ecol*. 19(10):2089–2104
- Attygalle AB, Wu X, Will KW (2006) Biosynthesis of Tiglic, Ethacrylic, and 2-Methylbutyric acids in a Carabid beetle, *Pterostichus (Hypherpes) californicus*. *J Chem Ecol* 33(5):963–970
- Attygalle AB, Xu S, Moore W, McManus R, Gill A, Will K (2020) Biosynthetic origin of benzoquinones in the explosive discharge of the bombardier beetle *Brachinus elongatulus*. *Sci Nat* 107:26
- Badawy AA (2017) Kynurenine pathway of tryptophan metabolism: regulatory and functional aspects. *Int J Tryptophan Res* 10:1–20
- Ben-Sahra I, Hoxhaj G, Ricoult SJH, Asara JM, Manning BD (2016) mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. *Science*. 351(6274):728–733

- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57(1):289–300
- Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 34:525–527
- Brosnan ME, Brosnan JT (2016) Formate: the neglected member of one-carbon metabolism. *Annu Rev Nutr* 36:369–388
- Brückner A, Parker J (2020) Molecular evolution of gland cell types and chemical interactions in animals in animals. *J. Exp. Biol* 223: jeb211938
- Carreras CW, Santi DV (1995) The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* 64:721–762
- Carroll N, Pangilinan F, Molloy AM, Troendle J, Mills JL, Kirke PN, Brody LC, Scott JM, Parle-McDermott A (2009) Analysis of the MTHFD1 promoter and risk of neural tube defects. *Hum Genet* 125(3):247–256
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552
- Chen S, Zhou Y, Taru C, Gu J (2018) Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890
- Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson SMW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol* 17:13
- Coppin CW, Jackson CJ, Sutherland T, Hart PJ, Devonshire AL, Russell RJ, Oakeshott JG (2012) Testing the evolvability of an insect carboxylesterase for the detoxification of synthetic pyrethroid insecticides. *Insect Biochem Mol Biol* 42(5):343–352
- Deuve T (1993) L'abdomen et les genitalia des femelles de Coleopteres Aephaga. *Mém Mus Natl Hist Nat* 155:1–184
- Di Pietro E, Sirois J, Tremblay ML, MacKenzie RE (2002) Mitochondrial NAD-dependent Methylene-tetrahydrofolate dehydrogenase-Methenyltetrahydrofolate Cyclohydrolase is essential for embryonic development. *Mol Cell Biol* 22(12):4158–4166
- Eddy SR (1998) Profile hidden Markov models. *Bioinformatics* 14(9):755–763
- Eisner T, Aneshansley D, del Campo ML, Eisner M, Frank JH, Deyrup M (2006) Effect of bombardier beetle spray on a wolf spider: repellency and leg autotomy. *Chemoecology* 16:185–189
- Eisner T, Meinwald J, Monro A, Ghent R (1961) Defence mechanisms of arthropods—I the composition and function of the spray of the whipscorpion, *Mastigoproctus giganteus* (Lucas) (Arachnida, Pedipalpida). *J Insect Physiol* 6(4):272–292
- Feyereisen R (2020) Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. *Mol Phylogenet Evol* 143:106695
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39(web server issue):W29–W37
- Forsyth DJ (1972) The structure of the pygidial defence glands of Carabidae (Coleoptera). *Trans Zool Soc Lond* 32:249–309
- Forsyth DJ (1968) The structure of the defence glands in the Dytiscidae, Noteridae, Halplidae and Gyrinidae (Coleoptera). *Trans R Ent Soc Lond* 120:158–181
- Forsyth DJ (1970) The structure of the defence glands of the Cicindelidae, Amphizoidae, and Hygrobiidae (Insecta: Coleoptera). *J Zool Lond* 160:51–69
- Fox JT, Stover PJ (2008) Folate-mediated one-carbon metabolism. *Vitam Horm* 79:1–44
- Francke W, Dettner K (2005) Chemical signalling in beetles. In: Schulz S (ed) *The chemistry of pheromones and other semiochemicals II*. Springer, Berlin, pp 85–166
- Froese DS, Fowler B, Baumgartner MR (2019) Vitamin B12, folate, and the methionine remethylation cycle- biochemistry, pathways, and regulation. *J Inherit Metab Dis* 42(4):673–685
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* 29(7):644–652
- Gustafsson R, Jemth AS, Gustafsson NM, Färnegårdh K, Loseva O, Wiita E, Bonagas N, Dahllund L, Llona-Minguez S, Häggblad M, Henriksson M, Andersson Y, Homan E, Helleday T, Stenmark P (2017) Crystal structure of the emerging Cancer target MTHFD2 in complex with a substrate-based inhibitor. *Cancer Res* 77(4):937–948
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512
- Han Q, Robinson H, Li J (2012) Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. *Biochem J* 446(2):253–260
- Haque MR, Higashiura A, Nakagawa A, Hirowatari A, Furuya S, Yamamoto K (2019) Molecular structure of a 5,10-methylenetetrahydrofolate dehydrogenase from the silkworm *Bombyx mori*. *FEBS Open Bio* 9(4):618–628
- Hartman SC, Buchanan JM (1959) Biosynthesis of the purines. XXVI The identification of the formyl donors of the transformylation reactions. *J Biol Chem* 234(7):1812–1816
- Hefetz A, Blum MS (1978a) Biosynthesis and accumulation of formic acid in the poison gland of the carpenter ant *Camponotus pennsylvanicus*. *Science* 201(4354):454–455
- Hefetz A, Blum MS (1978b) Biosynthesis of formic acid by the poison glands of formicine ants. *Biochim Biophys Acta* 543(4):484–496
- Ivarsson P, Henrikson BI, Stenson JAE (1996) Volatile substances in the pygidial secretion of gyrinid beetles (Coleoptera: Gyrinidae). *Chemoecology* 7(4):191–193
- Jackson CJ, Liu JW, Carr PD, Younus F, Coppin C, Meirelles T, Lethier M, Pandey G, Ollis DL, Russell RJ, Weik M, Oakeshott JG (2013) Structure and function of an insect α -carboxylesterase (α Esterase7) associated with insecticide resistance. *Proc Natl Acad Sci U S A* 110(25):10177–10182
- Jaffe JJ, Chrin LR (1979) De novo synthesis of methionine in normal and brugia-infected *Aedes aegypti*. *J Parasitol* 65(4):550–554
- Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P (2008) eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res* 36(Database issue):D250–D254
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14(6):587–589
- Kanai M, Funakoshi H, Takahashi H, Hayakawa T, Mizuno S, Matsumoto K, Nakamura T (2009) Tryptophan 2,3-dioxygenase is a key modulator of physiological neurogenesis and anxiety-related behavior in mice. *Mol Brain* 2:8
- Kanehisa K, Kawazu K (1982) Fatty acid components of the defensive substances in acid-secreting carabid beetles. *Appl Entomol Zool* 17(4):460–466
- Kanehisa K, Kawazu K (1985) Differences in neutral components of the defensive secretion in formic acid-secreting carabid beetles. *Appl Entomol Zool* 20(3):299–304
- Kanehisa K, Murase M (1977) Comparative study of the pygidial defensive systems of carabid beetles. *Appl Entomol Zool* 12:225–235
- Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res* 40:D109–D114

- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772–780
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3):567–580
- Kück P, Mayer C, Wägele JW, Misof B (2012) Long Branch effects distort maximum likelihood phylogenies in simulations despite selection of the correct model. *PLoS One* 7(5):e36593
- Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15:R29
- Lečić S, Čurčić S, Vujisić L, Čurčić B, Čurčić N, Nikolić Z, Anđelković B, Milosavljević S, Tešević V, Makarov S (2014) Defensive secretions in three ground-beetle species (Insecta: Coleoptera: Carabidae). *Ann Zool Fenn* 51:285–300
- Li T, White KP (2003) Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. *Dev Cell* 5(1):59–72
- Locher KP (2009) Structure and mechanism of ATP-binding cassette transporters. *Philos Trans R Soc Lond Ser B Biol Sci* 2364(1514):239–245
- Long HK, Prescott SL, Wysocka J (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 167(5):1170–1187
- MacFarlane AJ, Perry CA, Girmay HH, Gao D, Allen RH, Stabler SP, Shane B, Stover PJ (2009) Mthfd1 is an essential gene in mice and alters biomarkers of impaired one-carbon metabolism. *J Biol Chem* 284(3):1533–1539
- McCullough T (1966) Compounds found in the defensive scent glands of *Harpalus caliginosus* (Coleoptera: Carabidae). *Ann Entomol Soc Am* 59(5):1020–1021
- Mehler AH, Knox WE (1950) The conversion of tryptophan to kynurenine in liver. II The enzymatic hydrolysis of formylkynurenine. *J Biol Chem* 187(1):431–438
- Meiser J, Tumanov S, Maddocks O, Labuschagne CF, Athineos D, Van Den Broek N, Mackay GM, Gottlieb E, Blyth K, Voudsen K, Kamphorst JJ, Vazquez A (2016) Serine one-carbon catabolism with formate overflow. *Sci Adv* 2(10):e1601273
- Moore BP, Wallbank BE (1968) Chemical composition of the defensive secretion in carabid beetles and its importance as a taxonomic character. *Proc R Ent Soc Lond* 37(5–6):62–72
- Moore BP (1979) Chemical defense in Carabids and its bearing on phylogeny. *Carabid Beetles*:193–203
- Mulder GJ (1992) Glucuronidation and its role in regulation of biological activity of drugs. *Annu Rev Pharmacol Toxicol* 32:25–49
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268–274
- Parkhitko AA, Jouandin P, Mohr SE, Perrimon N (2019) Methionine metabolism and methyltransferases in the regulation of aging and lifespan extension across species. *Aging Cell* 18(6):e13034
- Peschke K, Eisner T (1987) Defensive secretion of the tenebrionid beetle *Blaps mucronata* physical and chemical determinants of effectiveness. *J Comp Physiol A* 161(3):377–388
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786
- Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F (2005) Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol* 5:50
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301
- Qiu Y, Tittiger C, Wicker-Thomas C, Le Goff G, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R (2012) An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A* 109(37):14858–14863
- Quistgaard EM, Löw C, Guettou F, Nordlund P (2016) Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat Rev Mol Cell Biol* 17:123–132
- Ranwez V, Harispe S, Delsuc F, Douzery EJP (2011) MACSE: multiple alignment of coding SEquences accounting for Frameshifts and stop codons. *PLoS One* 6(9):e22594
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47
- Rork AM, Mikó I, Renner T (2019) Pygidial glands of *Harpalus pensylvanicus* (Coleoptera: Carabidae) contain resilin-rich structures. *Arthropod Struct Dev* 44(12):1069–1083
- Rork AM, Renner T (2018) Carabidae Semiochemistry: current and future directions. *J Chem Ecol* 49:19–25
- Rossini C, Attygalle AB, González A, Smedley SR, Eisner M, Meinwald J, Eisner T (1977) Defensive production of formic acid (80%) by a carabid beetle (*Galerita lecontei*). *Proc Natl Acad Sci U S A* 94:6792–6797
- Roth LM, Eisner T (1962) Chemical defenses of arthropods. *Annu Rev Entomol* 7:107–136
- Savarse TM, Shih-Hsi C, Ming-Yu C, Parks RE (1985) 5'-Deoxy-5'-methylthioadenosine phosphorylase—III: role of the enzyme in the metabolism and action of 5'-halogenated adenosine analogs. *Biochem Pharmacol* 34(3):361–367
- Schildknecht H, Holoubek K (1961) Die Bombardierkäfer und ihre Explosionschemie V. Mitteilung über Insekten-Abwehrstoffe. *Angew Chem* 73(1):1–7
- Schildknecht H, Maschwitz U, Winkler H (1968a) Zur Evolution der Carabiden-Wehrdrüsensekrete. Über Arthropoden-Abwehrstoffe XXXII *Naturwissenschaften* 55:112–117
- Schildknecht H, Weis KH (1961) Die chemische natur des wehrsekretes von *Pseudophonus pubescens* und *Ps. griseus*. VIII. Mitteilung über insektenabwehrstoffe. *Z Naturforsch B* 16:361–363
- Schildknecht H, Winkler H, Maschwitz U (1968b) Vergleichend chemische Untersuchungen der Inhaltsstoffe der Pygidialwehrblasen von Carabiden. Über Arthropoden-Abwehrstoffe XXXI *Z Naturforsch B* 23:637–644
- Schildknecht H (1961) Über Insekten- und Pflanzenabwehrstoffe, ihre Isolierung und Aufklärung. *Angew Chem* 73(17–18):629
- Schildknecht H (1964) Defensive substances of the arthropods, their isolation and identification. *Angew Chem Int Ed* 3(2):73–82
- Schildknecht H (1970) The defensive chemistry of land and water beetles. *Angew Chem Int Ed* 9(1):1–9
- Schuler MA, Berenbaum MR (2013) Structure and function of cytochrome P450s in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. *J Chem Ecol* 39(9):1232–1245
- Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M, Barton GJ (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22(6):839–851
- Scott JG, Wen Z (2001) Cytochromes P450 of insects: the tip of the iceberg. *Pest Manag Sci* 57(10):958–967
- Sekowska A, Ashida H, Danchin A (2019) Revisiting the methionine salvage pathway and its paralogues. *Microb Biotechnol* 12(1):77–97
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212
- Smith JD, McManus KF, Fraser HB (2013) A novel test for selection on cis-regulatory elements reveals positive and negative selection acting on mammalian transcriptional enhancers. *Mol Biol Evol* 30(11):2509–2518

- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* 32(5):1342–1353
- Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, DeMeo DL, Quackenbush J, Glass K, Kuijjer ML (2017) Understanding tissue-specific gene regulation. *Cell Rep* 21(4):1077–1088
- Steuernagel L, Meckbach C, Heinrich F, Zeidler S, Schmitt AO, Gültas M (2019) Computational identification of tissue-specific transcription factor cooperation in ten cattle tissues. *PLoS One* 14(5):e021647
- Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc Series B* 64:479–498
- Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577
- Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12(10):692–702
- The Gene Ontology Consortium (2000) Gene ontology: tool for the unification of biology. *Nat Genet* 25:25–29
- Wagner W, Breksa AP 3rd, Monzingo AF, Appling DR, Robertus JD (2005) Kinetic and structural analysis of active site mutants of monofunctional NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase from *Saccharomyces cerevisiae*. *Biochemistry* 44(39):13163–13171
- West MG, Barlowe CK, Appling DR (1993) Cloning and characterization of the *Saccharomyces cerevisiae* gene encoding NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase. *J Biol Chem* 268(1):153–160
- Wheeler TJ, Eddy SR (2013) Nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29(19):2487–2489
- Will KW, Attygalle AB, Herath K (2000) New defensive chemical data for ground beetles (Coleoptera: Carabidae): interpretations in a phylogenetic framework. *Biol J Linn Soc* 71:459–481
- Will KW, Gill AS, Lee H, Attygalle AB (2010) Quantification and evidence for mechanically metered release of pygidial secretions in formic acid-producing carabid beetles. *J Insect Sci* 10:1–17
- Wilson EO, Regnier FE (1971) The evolution of the alarm-defense system in the formicine ants. *Am Nat* 105(943):279–289
- Wray GA (2007) The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* 8:206–216
- Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* 11:R14