The Pennsylvania State University

The Graduate School

DEFENSIVE CHEMICAL EVOLUTION IN THE CARABIDAE WITH AN EMPHASIS ON FORMIC ACID, METHACRYLIC ACID, AND BENZOQUINONE-PRODUCING TAXA

A Dissertation in

Entomology

by

Adam M. Rork

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2021

The dissertation of Adam M. Rork was reviewed and approved by the following:

Tanya Renner
Assistant Professor of Entomology
Dissertation Advisor
Chair of Committee

Etya Amsalem
Assistant Professor of Entomology

Julie Urban
Associate Research Professor of Entomology

Shaun Mahony
Associate Professor of Biochemistry & Molecular Biology

Gary Felton
Professor of Entomology
Head of the Department of Entomology

ABSTRACT

Of the approximately one million described insect species, ground beetles (Coleoptera: Carabidae) have long captivated the attention of evolutionary biologists due to their taxonomic diversity and array of defensive compounds secreted. Indeed, over 250 compounds have been documented from the defensive abdominal glands of ground beetles. This array of defensive strategies across the family make them ideal for understanding the evolution of insect chemical defense systems.

Despite their fascinating array of defensive compounds, there have been no investigations into the evolution of ground beetle pygidial glands. My overarching goal is to understand the underlying genetics, biochemistry, and evolution of ground beetle chemical defense as well as morphological adaptations of the glands to storing and secreting cytotoxic compounds. To do this, I am studying six ground beetles from three subfamilies: *Harpalus pensylvanicus* and *Platynus angustatus* are formic acid producers, *Pterostichus moestus* is a methacrylic acid producer, and the bombardier beetles *Brachinus elongatulus* and *Pheropsophus jessoensis* are benzoquinone producers from the subfamily Brachininae while *Goniotropis kuntzeni* is a benzoquinone producer from the subfamily Paussinae.

My research is driven by the following hypotheses: 1) Formic acid-producing ground beetles have co-opted the folate cycle of C1 metabolism for defensive formic acid biosynthesis. 2) Methacrylic acid-producing ground beetles have co-opted the valine catabolic pathway for defensive methacrylic acid biosynthesis. 3) Benzoquinone-producing ground beetles have co-opted a suite of oxidoreductase genes for the oxidation of phenolics, ultimately leading to benzoquinone biosynthesis. My research is also driven

by an exploratory interest in identifying potential morphological adaptations in the pygidial glands to defensive chemical resistance.

To address these hypotheses, RNA-Seq data were generated for each species' defensive glands and glandless bodies. Differential gene expression analyses were conducted between tissues, and significantly upregulated genes flagged as candidates for defensive chemical biosynthesis based on functional annotations. Isotopic precursors were also fed to select taxa and defensive chemicals harvested for gas chromatography-mass spectroscopy for confirmation of precursor incorporation.

Our results show that genes involved in the folate cycle of C1 metabolism are significantly upregulated in the defensive glands of formic acid producers compared to regular body tissue with isotope tracing results currently pending. Our results also support the hypothesis that L-valine catabolism plays a role in the biosynthesis of methacrylic acid and other related compounds in *P. moestus*. While we do find support for the presence of oxidoreductases in the secretory lobes of the three benzoquinone producers, it remains uncertain which if any are involved in benzoquinone biosynthesis. However, we have provided evidence for the role that peroxidases likely play in converting hydroquinones to their benzoquinone end-products, as well as genes involved in hydrogen peroxide biosynthesis in the Brachininae.

The glands of all species examined thus far, including *H. pensylvanicus*, *B. elongatulus*, and *Metrius contractus* (Paussinae) have been found to contain a high proportion of the elastomeric protein resilin in the collecting ducts, which is resistant to chemical and physical stress. By comprising the ductwork of the gland system, resilin presumably allows for chemicals to be transported to the reservoir chambers without

corroding the system. Generally found in arthropod body segments, this novel use of resilin in the gland system suggests it may be an adaptation to defensive chemical storage and transport.

My dissertation research thus provides novel insight into the evolution of insect defensive glands using the Carabidae as a case study. It is likely that the co-option of primary metabolic pathways through gene upregulation in the defensive glands is a key factor in the defensive chemical capabilities of ground beetles, and that morphological innovations such as resilin incorporation into gland tissue are important adaptations to defensive chemical storage and secretion.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation wouldn't have been possible without the support of my advisor, Dr. Tanya Renner. She gave me the freedom to study what most interested me about ground beetles, the encouragement that I could do anything I set my mind to, and the occasional reminder to take breaks when I needed them. Thank you for everything you've taught me, the opportunities you've provided me with, and for all your help over the years.

I also wish to thank my committee members, Etya Amsalem, Julie Urban, and Shaun Mahony for all their support and constructive feedback, from my comprehensive exam onwards. From chemical ecology to evolutionary biology to computational biology, their combined expertise, experience, and guidance have been of tremendous benefit in helping me improve my own research program and have undoubtably made me a better scientist in the process.

To István Mikó, thank you for introducing me to the fascinating world of insect morphology and comparative gland anatomy. Were it not for your passion, enthusiasm, and encouragement that I get to "Know my Insect", my research program would not be as well-rounded as it is today. To John Cantolina and Missy Hazen, thank you for teaching me how to use the wide array of instrumentation in the Microscopy Core Facility and for helping me troubleshoot my histological and microscopy protocols. To Andy Deans, thank you for the trust you instilled in me by allowing me to hold a key to the Frost Entomological Museum, thus allowing me to access specimens, space, and equipment in my quest to identify the carabids I've collected over the years.

I owe several thanks to all my collaborators, collectively known as the Carabid-Q Team, for the tremendous amount of support they've provided over the years. Thank you

to Athula Attygalle (SIT), Wendy Moore (UA), Kip Will (UC Berkeley), Aman Gill (UC Berkeley), Ramu Errabelli (SIT) Sihang Xu (SIT), Zhaoyu Zheng (SIT), Davide Bergamaschi (UA), Reilly McManus (UA), Sean Perez (UC Berkeley), Kylie Bocklund (PSU), Blue Riggins (PSU), and Arthi Bala (PSU).

I also owe great thanks to my undergraduate advisor Kyra Krakos, for giving me my first glimpse into biological research and for inspiring me to go to graduate school. Your mentorship, encouragement, and reminder to "be a good mammal" has helped me grow into the person and scientist I am today.

To all my wonderful friends near and far, thank you for the kindness and companionship you've shown me over the years. Whether it was inviting me over for dinner, our annual Friendsgiving, greeting me at coffee hour, or any other thoughtful act, all of it was sincerely appreciated. Thank you to Jonah Ulmer, Emily Sandall, Mariam Taleb, Erin Treanore, Brooke Lawrence, Laura Porturas, Codey Mathis, Shelby Kilpatrick, Avehi Singh, Stephania Sandoval Arango, Shelley Whitehead, Joyce Sakamoto, Chloe Drummond, Kadeem Gilbert, Blue Riggins, Yunjia Liu, Tristan Cofer, Hannah Tiffin, Fhallon Ware-Gillmore, Tyler Jones, and Stephanie Rodriguez to name only a few.

This work also would not have been possible without the generous support of the National Science Foundation and the Coleopterists Society.

To my parents, Kelly, Rick, Willie, and Jamie, to my siblings, Emily, Ashley, and Jacob, and to my entire family, thank you for the endless support you've shown me over the years and for always encouraging me to strive toward my goals.

EPIGRAPH

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim
Because it was grassy and wanted wear,
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has made all the difference.

Robert Frost, The Road Not Taken (1916)

DEDICATION

To my entire family, all my friends, and my mentors for believing in and supporting me every step of the way.

To all the frontline workers who have kept us safe and continue to do so during the COVID-19 pandemic.

CHAPTER 1

Carabidae Semiochemistry: Current and Future Directions

This work has been previously published and has been reproduced with permission from Springer Nature (https://doi.org/10.1007/s10886-018-1011-8).

ABSTRACT

Ground beetles (Carabidae) are recognized for their diverse, chemically-mediated defensive behaviors. Produced using a pair of pygidial glands, over 250 chemical constituents have been characterized across the family thus far, many of which are considered allomones. Over the past century, our knowledge of Carabidae exocrine chemistry has increased substantially, yet the role of these defensive compounds in mediating behavior other than repelling predators is largely unknown. It is also unclear whether non-defensive compounds produced by ground beetles mediate conspecific and heterospecific interactions, such as sex-aggregation pheromones or kairomones, respectively. Here we review the current state of non-exocrine Carabidae semiochemistry and behavioral research, discuss the importance of semiochemical research including but not limited to allomones, and describe next-generation methods for elucidating the underlying genetics and evolution of chemically-mediated behavior.

INTRODUCTION

Ground beetles (Carabidae) have long captured the attention of evolutionary biologists and chemical ecologists due to their great diversity and array of chemical defensive strategies (Darwin 1846, Eisner 1958). Since Thomas Eisner's pioneering studies on bombardier beetles, knowledge of carabid defensive chemistry has grown tremendously, with over 250 distinct chemical classes currently described (Lečić et al. 2014). Despite the importance of non-allomonic semiochemicals in mediating interactions between and within species, studies on behaviorally-active compounds for any of the over 40,000 species of carabid remain absent (Arndt et al. 2005). Given that many of these compounds are important, if not crucial, to regulating behavior in insects (e.g. sex-aggregation pheromones), understanding their biosynthesis and diversity could provide new insights into the evolution of one of the animal kingdom's most biodiverse families. We review the current state of non-allomonic semiochemistry and behavioral research in Carabidae and discuss new methods for studying the underlying genetics and evolution of biosynthetic pathways responsible for synthesis of these semiochemicals.

CARABIDAE SEMIOCHEMISTRY

Pygidial gland form and function

Pygidial glands are located posterodorsally within the abdomens of all adephagans, which include Carabidae (Forsyth 1968, 1970, 1972, Dettner 1985, Beutel & Hass 1996). Each independent glandular system is composed of one set of multi-lobed secretory glands, which are connected to reservoir-pump chambers via long collecting ducts (Forsyth 1972, Di Giglio et al. 2011). Defensive chemicals are thought to be synthesized in the multi-

lobed secretory glands, which funnel into the collecting duct, and ultimately collect in the reservoir chambers (Will et al. 2000, Attygalle et al. 2006). When a beetle is attacked, the pygidial glands are discharged by oozing the contents out of the tip of the abdomen, deterring the threat (Eisner et al. 1976, 2006). Within the Brachininae and Paussinae, however, an additional gross anatomical structure allows for the aiming and spraying of allomones at a target with great accuracy (Eisner & Aneshansley 1999). In these taxa, colloquially known as bombardier beetles, a highly-sclerotized, conical reaction chamber attaches to the posterior end of the reservoir and opens to the abdominal tip (Di Giulio et al. 2015, Eisner et al. 2000). Situated within the reaction chamber is a one-way valve that physically separates the contents of the reservoir chamber from the inside of the reaction chamber (Arndt et al. 2015). Accessory glands associated with the reaction chamber are thought to synthesize oxidative enzymes. When the valve is opened, hydroquinones and hydrogen peroxide rush into the reaction chamber, the hydroquinones are ultimately oxidized to p-benzoquinones, and the pressure build up ejects the mixture of newly oxidized quinone out of the abdomen at temperatures of up to 100°C at a rate of 341 to nearly 1000 pulses per second (Fig. 1) (Aneshansley et al. 1969, Eisner et al. 1977, Dean et al. 1990, Arndt et al. 2015). This mode of defense is considered apomorphic in these two clades with respect to the general strategy of oozing or light spraying found in most carabids (Di Giglio et al. 2011).

Defensive chemical diversity

Carabidae defensive compounds span 19 chemical classes and over 250 individual compounds, ranging from the simplest carboxylic acids to comparatively complex

benzoquinones and monoterpenes (Lečić et al. 2014). Carabids have been shown to produce saturated and unsaturated carboxylic acids, most notably formic, acetic, methacrylic, and tiglic acids (Schildknecht et al. 1968a, 1968b, Moore & Wallbank 1968, Eisner et al. 1977, Blum et al. 1981, Dettner 1987, Attygalle et al. 1991, Will et al. 2000, Francke & Dettner 2005, Lečić et al. 2014). The presence of these acids appears to be sexually dimorphic in some species, with female beetles producing the unsaturated compounds while males produce their fully saturated form (Attygalle et al. 1991). This is the only case of sexual dimorphism in Carabidae pygidial gland constituents that we are aware of. It is possible, albeit untested, that sexual dimorphism in pygidial compounds could play a role in mediating behavior. Formic acid is generally accompanied by the co-occurrence of unbranched hydrocarbons in the pygidial glands as well, as is the case for Formicine ants where they are potentially used as a surfactant to enhance the effect of their formic acid defensive sprays (Will et al. 2000). Some subfamilies produce salicylaldehyde and other phenolic derivatives, though this is seemingly less common than the production of organic acids. There have also been some reports of aromatic esters, aldehydes, aliphatic ketones, terpenes, hydrogen cyanide, and others in certain taxa (Blum et al. 1981, Schildknecht et al. 1968a, 1968b). One chemical class that is of particular interest are the hydroquinones and their oxidized quinone counterparts (Table 1). While quinones are present in four Carabidae subfamilies, Brachininae and Paussinae have evolved a unique way of utilizing them via rapid oxidation of hydroquinones to benzoquinones in a sclerotized reaction chamber (Schildknecht et al. 1968a, 1968b, Schildknecht 1970, Balestrazzi et al. 1985).

In all cases so far mentioned, these chemicals represent those found in the adult development stage. Only two studies have examined glandular secretions of larval and pupal stages. When examining the larvae of *Chlaenius cordicollis*, Holliday et al. 2015 discovered that a gland extruded in response to physical duress produced nine compounds, only one of which was found in the adult stage (2-methoxy-4-methylphenol). Six of these compounds, all quinones and phenolics, were suggested to be related to their adult counterparts by being their more oxidized versions, suggesting common biosynthetic pathways minus terminal reduction steps. 2-methoxy-4-ethylphenol was not suggested to be a precursor of any adult compounds by simple redox steps (Holliday et al. 2015). Di Giglio et al. 2009 discovered an enormous diversity of compounds in a single carabid species while examining the exocrine glands of *Carabus lefebvrei* pupae. In their glandular secretions, p-benzoquinone, two organic acids, numerous terpenes, aldehydes, and other compounds were detected; thirty-one compounds in total, two of which were isomers of β-ocimene (Di Giglio et al. 2009). Although these are only two studies, they exemplify how diverse carabid defensive metabolomes can be not only across taxa, but within taxa across developmental stages. It is possible that some of these compounds play roles in mediating behavior between members of the same species, especially within aggregations (Fig. 2).

Non-exocrine semiochemistry

Very few studies have examined carabid semiochemistry outside of the context of exocrine gland secretions. There are currently two accessible studies which examine cuticular chemistry, and there is only one that examines volatile emission from the entire

beetle. Bonacci et al. 2008 examined the cuticular chemistry of three carabid species: aposematic *Brachinus sclopeta* and *Anchomenus dorsalis*, as well as non-aposematic *Poecilus cupreus*, all of which can be found together in aggregations. It had been observed that *A. dorsalis* would exhibit a rubbing behavior toward *B. sclopeta*, which was initially suggested to be chemically-mediated and served as an odor-sharing mechanism (Brandmayr et al. 2006). *A. dorsalis* individuals were even observed to rub strips of paper that had been presented in monospecific *B. sclopeta* colonies, but not the control strips themselves. However, isolated individuals of *A. dorsalis* and *B. sclopeta* had very similar cuticular compound profiles compared to the non-aposematic co-aggregator, *Poecilus cupreus*. This suggested that most of the similarity seen was innate and due to the evolution of a Müllerian mimicry complex rather than odor sharing within an aggregation (Bonacci et al. 2008). However, it could be that the similarity in cuticular hydrocarbon profiles was acquired prior to individuals being separated for species-specific testing.

To investigate if cuticle chemistry is diet-dependent, Talarico et al. 2009 examined SPME extracts from a myrmecophilous beetle species, *Siagona europaea*, as well as one of their primary food sources, *Tapinoma nigerrimum*. It was discovered that beetles that had been fed ants resembled the cuticular chemical make-up of the ants more than unfed and isolated beetles. Fed beetles had acquired through some means an additional ten compounds compared to the fourteen found in unfed beetles, all of which were also present in the ant prey. Most of these compounds were aliphatic ketones, alkanes, and unsaturated hydrocarbons, as well as iridomyrmecin and isopulegol, although many remained unidentified. The same general trend appeared in beetles that preyed upon *Messor capitatus*; when fed this ant species, the beetles acquired a total of

sixteen compounds, doubling the eight that they had when isolated from ants. All identified compounds were alkanes, unsaturated hydrocarbons, as well as one ester. How these compounds are transferred from ant to beetle is currently unknown, although it is likely that they were acquired through physical contact rather than sequestration. (Talarico et al. 2009).

In contrast to the aforementioned non-volatile compounds characterized, Bonacci et al. 2011 was the first to examine the dynamic headspace volatiles released by carabid beetles, both placid and disturbed. In their analyses, they discovered that while each group produced only undecane, heneicosane, Z-9-tricosene, and tricosene, the disturbed beetles produced more of each compound, although only undecane was produced at a quantity that was found to be statistically significant (Bonacci et al. 2011).

POTENTIAL IMPLICATIONS OF CARABIDAE PHEROMONE & KAIROMONE STUDIES

Allomones as alarm pheromones

Carabid beetles produce an extraordinary diversity of exocrine compounds, many of which have been described as allomones (Schildknecht et al. 1968a, 1968b, Moore & Wallbank 1968, Will et al. 2000). While the efficiency of these sprays on predator deterrence has been assessed, the behavioral effects on the emitter and conspecifics have yet to be explored (Eisner & Dean 1976, Eisner et al. 2006), (Fig. 2). It has been demonstrated in other insect orders that volatiles released from sprays, stings, and other defensive behaviors can elicit alarm behavior in conspecifics, and this may hold true for carabids. The dual use of allomones as alarm pheromones is considered *parsimonious*

semiochemistry (Blum 1996). That is, it is suggested to be more parsimonious to use an honest signal of danger, an allomone, as an alarm pheromone than it is to have two separate compounds or blends of compounds for defense and alarm respectively. While this phenomenon of semiochemical bifunctionality is most often discussed in the context of eusocial insects, it has been also described in a number of solitary and gregarious species as well (Wilson & Regnier 1971, Farine et al. 2002, Cheng et al. 2017, Lockwood & Story 1987, Gunawardena & Bandumathie 1993, Machado et al. 2002, Löfqvist 1976, Blum 1969, 1996). Conspecific receivers of an alarm pheromone signal display one of two behavioral responses: (1) an increase in aggressive behavior toward the source of the threat, or (2) dispersal away from the source of the threat. These have been classically referred to as aggressive alarms and panic alarms, respectively (Wilson & Regnier 1971). In eusocial insects, the behavioral response may be caste specific, such as with the termite *Nasutitermes exitiosus*. Defensive sprays of *N. exitiosus* soldiers have been shown to recruit other soldiers to the emitter, while workers avoid the emission source if possible (Eisner et al. 1976, Costa-Leonardo & Haifig 2010). In the social wasp *Vespula velutina*, three aliphatic ketones released from the venom gland elicit an aggressive, stinging response in conspecifics (Cheng et al. 2017). In non-eusocial insects, such as the cockroach *Therea petiveriana* and the stink bug *Nezara viridula*, the seemingly universal response to these bifunctional defensive-alarm pheromones is dispersal (Farine et al. 2002, Lockwood & Story 1987). There is no documented behavior that we could find in the literature suggesting that non-social insects will risk predation to deter a threat when they themselves or their kin are not actively being attacked. That is, non-social insects are unlikely to risk predation to defend an unrelated conspecific. Thus, eusocial insects tend

to respond to defensive-alarm pheromones as “unselfish swarms”, their specific behavior being taxa, caste, and context dependent, while non-social insects behave as selfish herds (Hamilton 1971, Young et al. 1994). This is to be expected, given that these eusocial insects have the genetic incentive as haplodiploids to protect their reproductive caste, even at the risk of being preyed upon or having their siblings preyed upon. Solitary insects generally have no such incentive, and thus disperse to retain the opportunity for reproduction or protect their kin (Hamilton 1964a, 1964b, 1972, Trivers & Hare 1976).

Carabids offer an interesting opportunity to explore the role of defensive chemicals as alarm pheromones in a variety of ecological contexts (Fig. 2). While no carabid beetles are eusocial insects, certain taxa have been found to aggregate with conspecifics and, in some cases, heterospecifics under substrate (Bonacci et al. 2008, Philip & Burgess 2012, Thomas et al. 2001). Thus, carabids offer the unique opportunity to observe not only the response to defensive chemicals in solitary or gregarious species, but also between two to three different, co-aggregating species (Bonacci et al. 2008). This is especially true for the flanged bombardier beetles of the subfamily Paussinae, which are obligate or facultative myrmecophiles (Geiselhardt et al. 2007, Moore et al. 2011, Robertson & Moore 2016). It is currently unknown what behavioral effects their defensive blast of benzoquinones may have on their ant hosts, or what response formic acid or sting volatiles may elicit in the beetles. The behavioral effects that the defensive secretions of larvae and pupae may have on conspecifics is similarly a mystery.

Kairomone-mediated parasitism

Kairomones also play a significant role in mediating behaviors between insects, particularly between hosts, parasites, and parasitoids (Brown et al. 1970). Although parasitism on carabids by other organisms is not a topic that has been investigated in great detail, there are reports of wasp, mite, and nematode parasitism occurring in this group (Sasakawa et al. 2011, Fain et al. 2009, Andersen & Skorping 1990). Most interactions between parasites/parasitoids and their hosts are chemically mediated to some extent, the parasitic taxa orienting to chemical cues released by the host either directly or indirectly. For example, *Cotesia marginiventris* (Braconidae) utilize herbivore-induced plant volatiles to find the general location of their *Spodoptera frugiperda* hosts. Once on the plant, the wasp will then use chemical signatures left by the caterpillars on the leaf to locate their host and will parasitize them (Wölfling & Rostás 2009). Other parasitoids and parasites, such as the wasp *Aphytis melinus* and the nematode *Caenorhabditis japonica*, are attracted solely or primarily by volatile organic compounds released by the host itself rather than host-associated environmental signals (Stelicht 1973, Okumura & Yoshiga 2014). We know nothing, however, about the extent to which chemical eavesdropping on adult and immature carabid beetles by parasitoids and parasites may influence parasitic behavior (Fig. 2). One can only speculate on the diversity of parasites and parasitoids that afflict this hyper-diverse lineage, especially given that most preferentially infect one to a few species and may specialize on specific life stages (Strand & Obrycki 1996, Althoff 2003, Brodeur 2000).

Some carabid taxa are also reported as being ectoparasitoids on other organisms. Studies on North American *Brachinus* have shown that the larvae of some taxa are

parasitoids of aquatic beetle pupae, notably within the families Hydrophilidae, Gyrinidae, and Dytiscidae (Erwin 1967, Juliano 1984, Saska & Honek 2004). Since these pupae are nested within moist soil or under rocks obscuring them from view, there may be a chemical cue being released by the aquatic beetle pupae that is allowing the first-instar larvae or adult females to locate a host. Chemical eavesdropping on subterranean insects by surface-dwelling parasitoids has, to our knowledge, not been investigated before in insects. An investigation into the mechanisms regulating host-finding by *Brachinus* larvae could provide new insights into how parasitoids locate visibly hidden hosts against the noisy chemical background of the soil and the inhabitants therein.

Brachinus are not the only bombardier beetles to parasitize another species. Members of the subfamily Paussinae have obligate associations with ants and will prey upon them from within the colony (Moore et al. 2011, Robertson & Moore 2016). Their larvae employ this same predation tactic, neither seeming to evoke any usual aggressive response from their hosts (Geiselhardt et al. 2007). This is unusual behavior for ants to exhibit, a lack of aggressive response to a non-nest mate—even within a species, non-nest mates with the "incorrect" cuticular chemical profile are expelled or killed by worker or soldier castes (Martin et al. 2008, Sturgis & Gordon 2012, Emery & Tsutsui 2013, di Mauro et al. 2015). The ability of Paussine parasites to enter the colony and kill workers without evoking the colony's aggression has been explained partially by their potential ability to mimic ant stridulations (Di Giulio et al. 2014). There is likely a chemical component to the mediation of this behavior, although no specific compounds have been isolated either from the beetle cuticle or glandular exudates from which ants feed (Geiselhardt et al. 2007). That in itself, the beetles providing a food source for the ants,

could be enough to explain their protected status within the colony. Potentially, the beetles acquire cuticular compounds from constant contact with ants, or from consuming them, which would allow them to chemically blend in with their hosts (Elgar & Allan 2004, Johnson et al. 2001, Vander Meer & Wojcik 1982). They may also synthesize them independent of their hosts via convergence on a suite of biochemical pathways for producing cuticular constituents (Espelie & Hermann 1998, Allan et al. 2002, Geiselhardt et al. 2006). In some cases, myrmecophilous insects have been shown to release compounds that placate their ant hosts rather than attempting to blend in (Visicchio et al. 1999, Mori et al. 2000a, 2000b). However, this is generally a short-lived behavioral modification that would not well-explain the beetles' ability to remain in the nest for prolonged periods. It is also possible that they only possess a select few cuticular compounds that may be unimportant for nest mate recognition, essentially making themselves undetectable (Kleeberg et al. 2017). Other carabid taxa, such as larvae within the tribe Peleciini, are suggested to be ectoparasitoids of juvenile millipedes and Chrysomelid pupae (Erwin 1979, Salt 1928). The genus *Lebistina* are also parasitioids of Chrysomelid pupae, and are interestingly used by members of the San tribe of Africa to poison-tip arrows (Weber et al. 2008). While it is true that most carabid larvae are free living scavengers, the ectoparasitic taxa mentioned here could serve as interesting focal species for research into chemically-mediated parasite-host interactions.

Enhanced understanding of pheromone diversity and carabid evolution

Animals communicate through a variety of mediums, including sight, sound, touch, taste, and smell (Bando 1991, McDonald 1989, Jones & Teeling 2006, Sasahara et al. 2012,

Weber 1973, Hallem et al. 2006). In insects, pheromones are at the forefront of chemical communication between conspecifics (Hansson & Stensmyr 2011). Since the identification of the first insect sex pheromone, bombykol, by Adolf Butenandt in 1959, we have discovered behaviorally-active pheromone constituents from thousands of insect species, with more being discovered daily (Butenandt et al. 1959, Symonds & Elgar 2007). However, despite the intensive work on Coleoptera pheromone characterization, there is no published literature regarding knowledge of pheromones in Carabidae (Francke & Dettner 2005). Pheromones have been shown to be incredibly diverse in their biosynthetic and evolutionary origins, chemical structure, blend complexity, and behavioral effects; thus, an analysis of pheromones in this group would provide key new insights into all of these aspects of biochemical evolution (Tillman et al. 1999, Yew & Chung 2015, Francke & Dettner 2005).

While many classes of pheromones have been behaviorally characterized in the literature, many chemical ecologists and ethologists have focused on sex pheromones, in certain contexts referred to as sex-aggregation pheromones (Landolt 1997, Cardé 2014). These compounds, as the name implies, positively mediate aggregative reproductive behavior between two individuals of the same species. Differentially produced by members of one, sometimes both biological sexes of a species, sex pheromones attract conspecifics of the opposite sex of the emitter, who may in turn be releasing a different pheromone constituent or blend into the immediate headspace (Cardé & Baker 1984, Xu et al. 2017). If pheromone constituents are picked up by other conspecifics, aggregations may form due to a signaling cascade advertising reproductive opportunity (Landolt 1997). Thus, although they may never be directly detected by the olfactory receptors of

conspecifics of the same sex, they may indirectly play a role in aggregation formation via the attraction of pheromone producing members of the sex opposite the emitter, and so on. These can be distinguished from aggregation pheromones *sensu stricto* which mediate aggregative behavior without necessarily playing a direct role in reproductive behavior (Cardé 2014). Rather, many aggregation pheromones are suggested to play primary roles in communal feeding, such as with *Ips* and *Drosophila*, mate-finding being an indirect side-effect (Wertheim et al. 2005, Raffa 2011, Mast et al. 2014). Of course, microhabitats have a carrying capacity dictated by the resources available. Higher population densities also put those aggregating at a higher risk of acquiring disease or being parasitized if infected individuals are present, or being preyed upon (Dempster & Pollard 1981, Ioannou et al. 2009, Fuller et al. 2012). To ensure that the signal cascade produced by the positive feedback loop of pheromone release and attraction doesn't lead to overcrowding, some insects will release anti-aggregation pheromones. As their name implies, these function similarly to allomones, deterring in this case conspecifics (hence their classification of "pheromones") to maintain appropriate population densities. This particular pheromone-induced anti-aggregation effect has been particularly well-studied the bark beetles (Bakke 1981, Borden et al. 1987, Miller et al. 1995, 2005). Given that gregarious behavior is present in many carabid taxa, whether it be for increased fecundity, feeding, defense, etc., it is conceivable that pheromones may be playing a significant role in the formation and maintenance of such aggregations (Alatalo & Mappes 1996, Gamberale & Tullberg 1998). While they would etymologically not be classified as pheromones given the context, there may be a chemical basis to the mediation of interspecific aggregation formation and maintenance as well (Walgenbach

et al. 1982). Given the potential importance of these aggregations during periods of inactivity and their relevance to the natural history of this group, we suggest that further investigation into these gregarious behaviors are warranted (Arndt et al. 2005).

When sex pheromones, whether they be blends or single constituents, are released at the proper concentrations, ratios, and time, they may serve as the lynch-pin for ensuring reproductive success. If there is even the slightest phenotypic variation in any of these factors, especially in the chemical make-up of a blend, the signal and expected behavior become desynchronized (Baker 1998, Kárpáti et al. 2013, Klun et al. 1973). That is, an inappropriate signal will not likely elicit the preferred response, such as attraction for mating. If this desynchronization is prevalent in a great enough proportion of the population in question, and if this truly does cause ethological reproductive isolation, speciation events may begin to occur (Wicker-Thomas 2009, Nanda & Singh 2012). These alterations to wild-type pheromone blends are generally the result of altered enzymatic activity, whether it be in terms of the quantity of pheromone constituents produced, their relative ratios, or chemical identity (Buček et al. 2015, Albre et al. 2013). In many Lepidoptera, primarily moths, most sex pheromones are synthesized from fatty acid precursors through a series of β-oxidation and desaturation reactions (Roelofs & Bjostad 1984). They may then be modified via acetylation, carbonyl reduction, or other biochemical steps to form the final pheromone product (Roelofs & Rooney 2003, Bjostad & Roelofs 1983). In the European corn borer, *Ostrinia nubialis*, two distinct chemotypes are present throughout the Midwestern and Northeastern United States. The Midwestern population primarily produce a 97:3 blend of Z:E-11-tetradecenyl acetate (TDA), while the Northeastern population produce a 3:97 blend of Z:E-11-TDA. Behavioral assays

show that male moths of one chemotype are often not significantly attracted to pheromone blends of the other (i.e. New York *Ostrinia* are seldom attracted to the pheromone blends of Iowan moths, and vice versa). However, in regions where the ranges of both chemotypes overlap (e.g. Pennsylvania), males were attracted to a 1:1 blend of the two pheromone isomers as well as the 97:3 and 3:97 blends (Klun et al. 1975, Kocbansky et al. 1975, Roelofs et al. 1985). It is conceivable that sustained ethological reproductive isolation over a long enough period of time may lead to the buildup of genetic differences such that speciation events occur (Jennings et al. 2011, 2014). Much of the early work on pheromone evolution was also conducted in the genus *Ostrinia,* particularly using *O. nubialis* and *O. furnacalis*, the European and Asian corn borers respectively. Both species produce structurally similar primary pheromone constituents, Z/E-11-TDA and Z/E-12-TDA respectively; a desaturation shift on either side of the twelfth carbon of the alkyl chain. In both *Ostrinia* species, Δ9, Δ11, and Δ14 desaturases were detected in the abdominal pheromone production glands. However, only Δ11-produced pheromone product (Z/E-11-TDA) was found in *O. nubialis* whereas *O. furnacalis* produces the Δ14 pheromone product (Z/E-12-TDA). Thus, somewhere in the evolutionary history of *Ostrinia*, a switch to the utilization of Δ14 occurred, leading to the observed chemotype of *O. furnacalis* (Roelofs et al. 2002). It was hypothesized that this may have occurred via the resurrection of a Δ14 pseudogene combined with some suppression of Δ11-desaturase transcription or enzymatic activity in this lineage. Similar examples of pheromone evolution and the role that it can play in the process of organismal evolution have been identified in other insects (Baker 2002, Symonds & Elgar 2003, Ferveur 2005). If studies into the genetic and biosynthetic bases of

pheromone evolution are undertaken in carabids, along with work into the neurophysiology and pheromone-mediated behavior of this group, we could gain valuable insights into how this group has radiated to the extent that is has should pheromone evolution be playing a significant role.

The Carabidae are one of the single largest families of insects currently known (Arndt et al. 2005). Members of this family have colonized nearly every habitat on land, and in rare cases, some have become semi-aquatic, such as the Fairy Shrimp Hunting Beetle, *Cicindis horni* (Erwin & Aschero 2004). One fossilized species has even been found in Antarctica from the early to mid-Miocene (14-20 Ma), and many extant species inhabit the arctic tundra (Ashworth & Erwin 2016, Garry 1993). Such biodiversity has captivated the attention of biologists and numerous mechanisms have been suggested to play a role in speciation. One such mechanism is reproductive isolation, be is allopatric or sympatric (Jennings et al. 2011, 2014, Bolnick & Fitzpatrick 2007). Given that pheromone evolution has been shown to be saltational in some cases, sympatric reproductive isolation via a sudden desynchronization in pheromone emission, olfaction, and behavior is one way in by which speciation events can occur (Baker 2002, Symonds & Elgar 2003, Ferveur 2005, Wicker-Thomas 2009, Nanda & Singh 2012). A broad investigation into the pheromone constituent of this family, their biosynthesis pathways, and the rates of evolution of the enzymes involved is necessary to make any inferences as to what role saltational shifts in pheromone blends may have had in the diversification of this lineage of beetles.

It is possible that we have unknowingly detected pheromone components in the defensive secretions and sprays of the carabid species examined thus far (Lečić et al.

2014). As previously mentioned, many defensive allomones may also serve roles as alarm pheromones, being that they are an honest signal of danger (Blum 1969, 1996, Eisner et al. 1976, Löfqvist 1976). However, in some cases, allomones may serve a dual role as sex pheromones. In the parasitoid wasp, *Leptopilina heterotoma*, the defensive compound (-)-iridomyrmecin functions not only as a defensive allomone, but also as a competition avoidance cue in the context of reproduction (Weiss et al. 2013). In the presence of other minor constituents such as (+)-isoiridomyrmecin and other structurally related compounds, it has also been found to function as a sex pheromone constituent (Weiss et al. 2013). Although this type of semiochemical parsimony has not been widely reported on in the literature, it may be more prevalent than we are currently aware (Ruther et al. 2001, Geiselhardt et al. 2008, Weiss et al. 2013). Given the diversity of natural histories, as well as the diversity of their defensive allomones, it may be carabids also employ semiochemical parsimony (Fig. 2). A natural first target for studies into alarm pheromone emission, detection, and chemically-mediated alarm behavior within the Carabidae would be primary pygidial gland defensive secretions. Not only are these compounds easily extractable from the reservoirs, most are also commercially available or easily synthesizable, such as benzaldehyde produced by members of the Cicindelinae (tiger beetles). An added benefit to utilizing the Carabidae as a model for studying alarm pheromone behavior is that the gland constituents are relatively simple within a species, most taxa usually only containing one to a few major compounds and occasionally a small number of minor constituents. While the usual challenges of running successful behavioral assays on insects would persist, many steps related to the identification of

multiple compounds produced in trace quantities would be eliminated due to the high quantity and relative homogeneity of intraspecific gland constituents in certain taxa.

NEXT-GENERATION CHEMICAL ECOLOGY

Over the past 60 years, our ability to identify insect semiochemicals has greatly improved. More than 3,500 unique compounds have been discovered in thousands of taxa, belonging to nearly every insect order (El-Sayed 2018). Many of these compounds have been assigned semiochemical classes and are hypothesized to have specific influences on insect behavior based on the ecological conditions of their emission. In contrast, the underlying genetics of semiochemical biosynthesis in many insects remain understudied, which is reflected by a lack of genetic data for many insect groups including carabids. Now, with the advent of next-generation sequencing technologies, many of the financial and logistical barriers that previously impeded the high-throughput study of genes involved in insect semiochemical biosynthesis have been greatly diminished. Studies can be conducted to identify genes involved in physiological processes for whole insect bodies, specific glandular tissues, and even individual cells (Ziegenhain et al. 2017), allowing for high-resolution analyses of differential gene expression between or within tissues. Integrating next-generation sequencing with analytical chemistry and biochemical techniques can allow for an enhanced understanding of semiochemical production and evolution throughout Insecta. To advance understanding of carabid semiochemistry specifically, we suggest an integrative approach that combines tissue-specific transcriptomics with spectroscopic and biochemical assays to elucidate complete biosynthetic pathway(s) for semiochemicals.

Next-generation sequencing technologies have revolutionized the method by which gene expression can be studied within an organism, tissue or cell. In particular, whole transcriptome sequencing, also known as RNA sequencing (RNA-Seq), can be used to analyze context-dependent gene expression utilizing billions of sequenced base pairs (Conesa et al. 2016). For example, transcript abundance can be used to identify genes that are differentially expressed between samples (e.g. tissue of interest relative to the whole body) and have putative functions in tissue-specific biosynthetic processes. RNA-Seq has already proven to be a successful method for the identification of candidate genes involved in pheromone biosynthesis (Vogel et al. 2010; Buček et al. 2015, 2016; Nadeau et al. 2017). For instance, transcriptome sequencing of the *Heliothis virescens* pheromone gland revealed high expression of several different candidate genes, which may play roles in pheromone biosynthesis (Vogel et al. 2010). Many of these were annotated as alcohol and aldehyde oxidases, fatty acid synthetases and reductases, and various fatty acid desaturases, which corroborates findings from previous studies on the biosynthesis of noctuid pheromone constituents that these genes are involved in pheromone biosynthesis (Teal & Tumlinson 1986, 1988). Similarly, RNA-Seq may be a promising tool for the identification of genes expressed in carabid tissues that biosynthesize semiochemicals, which could then serve as candidates in studies investigating biosynthesis (Fig. 3).

While RNA-Seq can provide support for tissue-specific patterns of gene expression, this method cannot confirm which transcripts are functionally relevant for a given metabolic process (Fig. 3). Two promising techniques in studying the role of gene products in the biosynthesis of semiochemicals are RNA interference (RNAi) and

CRISPR/Cas9, which can be used to knock down or knock out gene expression (Fire et al. 1998, Cong et al. 2013). The result is an observable phenotype, such as the absence or reduced production of a chemical constituent observable via gas chromatography–mass spectrometry (Li et al. 2013). Thus, from a set of differentially expressed genes, possible precursors, cofactors, and potentially even intermediates in the pathway can be inferred at least in-so-far as it narrows down the list of candidates involved. However, with CRISPR/Cas9, it is necessary for injections to be made in the egg stage of the insect, and thus a well-maintained colony is required (Kistler et al. 2015). Where this is unrealistic due to the complex life history strategies of many insects and difficulty of rearing, RNAi may serve as an alternative for assessing gene function, as it can be used at various life stages once optimized.

Radiolabeled precursor incorporation followed by candidate enzyme characterization should be combined with gene knockdown and knockout methods to assess whether genes expressed within carabid pygidial glands play a role in converting chemical precursors to semiochemical constituents. A prime example of such *in vivo* radiolabeled assays comes from Attygalle et al. 1991, in which the injection of D_8-L-valine into the carabid beetle *Scarites subterraneus* was shown to be highly incorporated into methacrylic acid and isobutyric acid, two pygidial gland allomones, via GC-MS analysis of gland extracts (Attygalle et al. 1991). An alternative to injections is the incorporation of radiolabeled precursors into the diet, which is particularly useful if the insect is too small to handle injections or if the stress and physical damage induced by a needle is undesirable. This method was historically popular in studies examining the essential amino acid requirements of insects (Kastings & McGinnis 1956, Rock & King

1966). The same method of feeding an insect radiolabeled candidate precursors, preferably those more immediately upstream of the compound(s) of interest, could also reveal which if any are incorporated into, for example, a carabid semiochemical constituent.

To validate the role of candidate genes in carabid semiochemical biosynthesis, protein expression followed by enzymatic assays can be performed on individual steps in biosynthetic pathways. For example, expression vectors can be designed from the candidate genes discovered via RNA-Seq for their introduction into cell lines (Ohlen et al. 2016, Steiner et al. 2018, Zhang et al. 2018). Once the purified protein isolate has been acquired, physiological conditions can be replicated by adding appropriate concentrations of relevant cofactors and the substrate of interest (Ohlen et al. 2016, Steiner et al. 2018, Zhang et al. 2018). The reaction can essentially be performed *in vitro*, and the conversion of substrate to product can be confirmed using GC-MS analysis. By conducting such assays, candidate genes identified in RNA-Seq experiments can be functionally validated via the confirmation of their role in semiochemical biosynthesis at the enzyme-substrate level.

To gain a deeper understanding of the processes by which semiochemical biosynthetic pathways evolve and diversify, the evolutionary history of the gene families involved should also be considered. A previous study of Lepidopteran fatty acid desaturases, which are important genes for pheromone biosynthesis in this order, serves as a prime example of how tissue-specific RNA-Seq can be combined with phylogenetics and molecular evolutionary analyses. Phylogenetic reconstruction of the Lepidopteran fatty acid desaturase gene family revealed interesting implications for pheromone

evolution in moths, including the existence of many functionally conserved, yet inactive clades that lie dormant within moth genomes (Buček et al. 2015). In addition, single point mutations were suggested to alter the substrate specificity of two paralogous enzymes, a Z11-desaturase/conjugase and an E/Z-14 desaturase in *Manduca sexta* (Buček et al. 2015). This shift in specificity via the mutation of a small number of amino acids was suggested to be a potential mechanism by which novel pheromone constituents could evolve. Similar studies in Carabidae have the potential to reveal interesting gene birth, death, and functional divergence events associated with shifts in semiochemical production (Fig. 3).

CONCLUSIONS

The exocrine chemistry of Carabidae have been investigated heavily over the past 60 years, and thanks to the efforts of many dedicated carabidologists and chemical ecologists, we have amassed a rich body of allomone literature across the family (Schildknecht et al. 1968a, 1968b, Moore & Wallbank 1968, Will et al. 2000). As important as this is to understanding the evolution of chemical defensive strategies, it could be complimented by studies of other chemically-mediated methods of intraspecific and interspecific communication. Without knowledge of how carabids mediate their complex interspecific aggregations, find hosts for predation deep within volatile-rich soil, survive the selective chemical screening by ant colonies, and ultimately find their mates in a complex world of chemical and visual cues, we leave much unknown about this ecologically important and diverse lineage. In addition, understanding the biosynthesis of carabid semiochemicals and the evolution of the complex behaviors that they mediate is

arguably as necessary as understanding the chemicals and behaviors themselves. Few studies have elucidated the underlying genetics of semiochemical biosynthetic pathways in insects, especially within Carabidae. Therefore, we strongly support expanding studies of carabid semiochemistry, olfaction, and behavior.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (DEB 1556931/1762760 to Tanya Renner). The findings and conclusions of this work do not necessarily reflect the views of the National Science Foundation. The authors thank Wendy Moore for comments on an earlier version of the manuscript, as well as Kipling Will and Athula Attygalle for thoughtful discussion and advice.

REFERENCES

1. Alatalo RV and Mappes J. (1996). Tracking the evolution of warning signals. *Nature*. 382(6593): 708–710.
2. Albre J, Liénard MA, Sirey TM, Schmidt S, Tooman LK, Carraher C, Greenwood DR, Löfstedt C, and Newcomb RD. (2012). Sex Pheromone Evolution Is Associated with Differential Regulation of the Same Desaturase Gene in Two Genera of Leafroller Moths. *PLOS Genetics.* 8(1): e1002489.
3. Allan RA, Capon RJ, Brown V, and Elgar MA. (2002). Mimicry of Host Cuticular Hydrocarbons by Salticid Spider *Cosmophasis bitaeniata* That Preys on Larvae of Tree Ants *Oecophylla smaragdina*. *Journal of Chemical Ecology.* 28(4): 835-848.
4. Althoff DM. (2003). Does parasitoid attack strategy influence host specificity? A test with New World braconids. *Ecological Entomology.* 28(4): 500-502.
5. Andersen J and Skorping A. (1990). Parasites of carabid beetles: prevalence depends on habitat selection of the host. *Canadian Journal of Zoology.* 69(5): 1216-1220.
6. Aneshansley DJ, Eisner T, Widom JM, and Widom B. (1969). Biochemistry at 100 °C: Explosive Secretory Discharge of Bombardier Beetles (Brachinus). *Science*. 165(3888): 61-63.
7. Arndt E, Beutel RG, and Will K. (2005). 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)*. (pp. 119–144). Walter de Gruyter, Berlin.

8. Arndt EM, Moore W, Lee WK, and Ortiz C. (2015). Mechanistic origins of the bombardier beetle (Brachinini) explosive pulsed chemical defense. *Science.* 348(6234): 563-567.
9. Ashworth AC and Erwin TL. (2016). *Antarctotrechus balli* sp. n. (Carabidae, Trechini): the first ground beetle from Antarctica. *ZooKeys*. 635: 109-122.
10. Attygalle AB, Meinwald J, Liebherr JK, and Eisner T. (1991). Sexual dimorphism in the defensive secretion of a carabid beetle. *Experientia*. 47(3): 296-299.
11. Attygalle AB, Wu X, and Will KW. (2006). Biosynthesis of Tiglic, Ethacrylic, and 2-Methylbutyric Acids in a Carabid Beetle, *Pterostichus* (*Hypherpes*) *californicus*. *Journal of Chemical Ecology.* 33(5): 963-970.
12. Baker TC. (1998). Species specificity of pheromone responses. *Biochemist.* 20: 26-29.
13. Baker TC. (2002). Mechanism for saltational shifts in pheromone communication systems. *Proceedings of the National Academy of Sciences of the United States of America.* 99(21): 13368-13370.
14. Bakke A. (1981). Inhibition of the response in *Ips typographus* to the aggregation pheromone; field evaluation of verbenone and ipsenol. *Journal of Applied Entomology*. 92(1-5): 172-177.
15. Balestrazzi E, Dazzini MLV, De Bernardi M, Vidari G, Vita-Finzi P, and Mellerio G. (1985). Morphological and chemical studies on the pygidial defence glands of some Carabidae (Coleoptera). *Naturwissenschaften.* 72(9): 482–484.
16. Bando T. (1991). Visual perception of texture in aggressive behavior of *Betta splendens*. *Journal of Comparative Physiology A.* 169(1): 51-58.

17. Beutel RG and Haas A. (1996). Phylogenetic analysis of larval and adult characters of Adephaga (Coleoptera) using cladistics computer programs. *Insect Systematics & Evolution*. 27(2): 197-205.

18. Bjostad LB and Roelofs WL. (1983). Sex pheromone biosynthesis is *Trichoplusia ni*: Key steps involve delta-11 desaturation and chain-shortening. *Science*. 220(4604): 1387-1389.

19. Blum MS, Jones TH, House GJ, and Tschinkel WR. (1981). Defensive secretions of tiger beetles: Cyanogenetic basis. *Comparative Biochemistry & Physiology.* 69(4): 903-904.

20. Blum MS. (1969). Alarm Pheromones. *Annual Review of Entomology.* 14: 57-80.

21. Blum MS. (1996). Semiochemical parsimony in the Arthropoda. *Annual Review of Entomology*. 41: 353-374.

22. Bolnick DI and Fitzpatrick BM. (2007). Sympatric Speciation: Models and Empirical Evidence. *Annual Review of Ecology, Evolution, and Systematics.* 38: 459-487.

23. Bonacci T, Brandmayr P, Dalpozzo R, De Nino A, Massolo A, Tagarelli A, and Brandmayr TZ. (2008). Odour and colour similarity in two species of gregarious carabid beetles (Coleoptera) from the Crati Valley, southern Italy: a case of Müllerian mimicry? *Entomological News*. 119(4): 325-337.

24. Bonacci T, Brandmayr P, Zetto T, Perrotta ID, Guarino S, Peri E, and Colazza S. (2011). Volatile compounds released by disturbed and undisturbed adults of *Anchomenus dorsalis* (Coleoptera, Carabidae, Platynini) and structure of the pygidial gland. *ZooKeys*. 81: 13-25.

25. Borden JH, Ryker LC, Chong LJ, Pierce HD, Johnston BD, and Oehlschlager, AC. (1987). Response of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Coleoptera: Scolytidae), to five semiochemicals in British Columbia lodgepole pine forests. *Canadian Journal of Forest Research*. 17(2): 118-128.
26. Brandmayr TZ, Bonacci T, Massolo A, and Brandmayr P. (2006). What is going on between aposematic carabid beetles? The case of Anchomenus dorsalis (Pontoppidan 1763) and Brachinus sclopeta (Fabricius 1792) (Coleoptera Carabidae). *Ethology, Ecology, & Evolution*. 4: 335-348.
27. Brodeur J. (2000). Host Specificity and Trophic Relationships of Hyperparasitoids. In Hochberg ME, Ives AR (Eds.), *Parasitoid Population Biology.* (pp. 163-183). Princeton, NJ: Princeton University Press.
28. Brown WL, Eisner T, and Whittaker RH. (1970). Allomones and kairomones: Transpecific chemical messengers. *BioScience*. 20(1): 21.
29. Buček A, Brabcová J, Vogel H, Prchalová D, Kindl J, Valterová I, and Pichová I. (2016). Exploring complex pheromone biosynthetic processes in the bumblebee male labial gland by RNA sequencing. *Insect Molecular Biology*. 25(3): 295-314.
30. Buček A, Matouškova P, Vogel H, Šebesta P, Jahn U, Weißflog J, Svatoš A, and Pichová I. (2015). Evolution of moth sex pheromone composition by a single amino acid substitution in a fatty acid desaturase. *Proc. Natl. Acad. Sci.* 112(41): 12586-12591.
31. Butenandt VA, Beckmann R, Stamm D, and Hecker E. (1959). Über den Sexual-Lockstoff des Seidenspinners *Bombyx mori* – Reindarstellung und Konstitution. *Zeitschrift für Naturforschung.* 14: 283–284.

32. Cardé RT and Baker TC. (1984). Sexual Communication with Pheromones. In: Bell WJ, Cardé RT (Eds.) *Chemical Ecology of Insects*. Springer, Boston, MA.

33. Cardé RT. (2014). Defining Attraction and Aggregation Pheromones: Teleological Versus Functional Perspectives. *Journal of Chemical Ecology.* 40(6): 519-520.

34. Cheng Y, Wen P, Dong S, Tan K, Nieh JC. (2017). Poison and alarm: the Asian hornet *Vespa velutina* uses sting venom volatiles as an alarm pheromone. *Journal of Experimental Biology.* 220: 645-651.

35. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, and Mortazavi A. (2016). A survey of best practices for RNA-Seq data analysis. *Genome Biology*. 17(13).

36. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, and Zhang F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science.* 339(6121): 819–823.

37. Costa-Leonardo AM and Haifig I. (2010). Pheromones and exocrine glands in Isoptera. *Vitamins & Hormones.* 83: 521-549.

38. Darwin Correspondence Project, "Letter no. 1009," accessed on 10 September 2017, http://www.darwinproject.ac.uk/DCP-LETT-1009.

39. Dean J, Aneshansley DJ, Edgerton HE, and Eisner T. (1990). Defensive spray of the bombardier beetle: a biological pulse jet. *Science*. 248(4960): 1219-1221.

40. Dempster JP and Pollard E. (1981). Fluctuations in Resource Availability and Insect Populations. *Oecologia*. 50(3): 412-416.

41. Dettner K. (1985). Ecological and phylogenetic significance of defensive compounds from pygidial glands of Hydradephaga (Coleoptera). *Proceedings of the Academy of Natural Sciences of Philadelphia*. 137(1): 156-171

42. Dettner K. (1987). Chemosystematics and evolution of beetle chemical defenses. *Annual Review of Entomology.* 32: 17-48.

43. Di Giglio A, Brandmayr P, Dalpozzo R, Sindona G, Tagarelli A, Talarico F, Brandmayr TZ, Ferrero EA. (2009). The Defensive Secretion of *Carabus lefebvrei* Dejean 1826 Pupa (Coleoptera, Carabidae): Gland Ultrastructure and Chemical Identification. *Microscopy Research and Technique.* 72(5): 351-361.

44. Di Giglio A, Brandmayr P, Talarico F, and Brandmayr TZ. (2011). Current knowledge on exocrine glands in carabid beetles: Structure, function, and chemical compounds. *ZooKeys*. 100: 193-201.

45. Di Giulio A, Fattorini S, Moore W, Robertson J, and Maurizi E. (2014). Form, function and evolutionary significance of stridulatory organs in ant nest beetles (Coleoptera: Carabidae: Paussini). *European Journal of Entomology.* 111(5): 692-702.

46. Di Giulio A, Muzzi M, and Romani R. (2015). Functional anatomy of the explosive defensive system of bombardier beetles (Coleoptera, Carabidae, Brachininae). *Arthropod Structure and Development.* 44(5): 468-490.

47. Di Mauro G, Perez M, Lorenzi MC, Guerrieri FJ, Millar JG, d'Ettorre P. (2015). Ants Discriminate Between Different Hydrocarbon Concentrations. *Frontiers in Ecology and Evolution*. 3: 133.

48. Eisner T and Aneshansley DJ. (1999). Spray aiming in the bombardier beetle: Photographic evidence. *Proceedings of the National Academy of Sciences of the United States of America.* 96(17): 9705-9709.

49. Eisner T and Dean J. (1976). Ploy and counterploy in predator-prey interactions: Orb-weaving spiders versus bombardier beetles. *Proceedings of the National Academy of Sciences of the United States of America.* 73(4): 1365-1367.

50. Eisner T, Aneshansley D, del Campo ML, Eisner M, Frank JH, and Deyrup M. (2006). Effect of bombardier beetle spray on a wolf spider: repellency and leg autotomy. *Chemoecology*. 16(4): 185-189.

51. Eisner T, Aneshansley DJ, Eisner M, Attygalle AB, Alsop DW, and Meinwald J. (2000). Spray mechanism of the most primitive bombardier beetle (*Metrius contractus*). *Journal of Experimental Biology.* 203: 1265-1275.

52. Eisner T, Jones TH, Aneshansley DJ, Tschinkel WR, Silberglied RE, and Meinwald J. (1977). Chemistry of Defensive Secretions of Bombardier Beetles (Brachinini, Metriini, Oxaenini, Paussini). *Journal of Insect Physiology.* 23(11-12): 1383-1386.

53. Eisner T, Kriston I, and Aneshansley DJ. (1976). Defensive Behavior of a Termite (*Nasutitermes exitiosus*). *Behavioral Ecology and Sociobiology.* 1(1): 83-125.

54. Eisner T. (1958). The protective role of the spray mechanism of the bombardier beetle, *Brachynus ballistarius* Lec. *Journal of Insect Physiology* 2(3): 215-220.

55. El-Sayed AM. (2018). The Pherobase: Database of Pheromones and Semiochemicals. http://www.pherobase.com.

56. Elgar MA and Allan RA. (2004). Predatory spider mimics acquire colony-specific cuticular hydrocarbons from their ant model prey. *Naturwissenschaften*. 91(3): 143-147.

57. Emery VJ and Tsutsui ND. (2013). Recognition in a Social Symbiosis: Chemical Phenotypes and Nestmate Recognition Behaviors of Neotropical Parabiotic Ants. *PLOS One*. 8(2): e56492.

58. Erwin TL and Aschero V. (2004). *Cicindis horni* Bruch (Coleoptera: Carabidae, Cicindini): The Fairy Shrimp Hunting Beetle, its way of life on the Salinas Grandes of Argentina. *Zootaxa*. 553: 1-16.

59. Erwin TL. (1967). Bombardier beetles (Coleoptera: Carabidae) of North America: Part II. Biology and behavior of *Brachinus pallidus* Erwin in Carlifonia. *Coleopterists Bulletin.* 21: 41-55

60. Erwin TL. (1979). A Review of the Natural History and Evolution of Ectoparasitoid Relationships in Carabid Beetles. In: Erwin T.L., Ball G.E., Whitehead D.R., Halpern A.L. (eds.) *Carabid Beetles*. Springer, Dordrecht.

61. Espelie KE and Hermann HR. (1988). Congruent cuticular hydrocarbons: Biochemical convergence of a social wasp, an ant and a host plant. *Biochemical Systematics and Ecology.* 16(5): 505-508.

62. Fain A, Noti MI, and Dufrêne M. (2009). Observations on the mites (Acari) associated with Carabidae (Coleoptera) in Belgium. I. Annotated list of the species. *International Journal of Acarology.* 21(2): 107-122.

63. Farine JP, Semon E, Everaerts C, Abed D, Grandcolas P, and Brossut R. (2002). Defensive secretion of *Therea petiveriana*: chemical identification and evidence of an alarm function. *Journal of Chemical Ecology.* 28(8): 1629-1640.

64. Ferveur JF. (2005). Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. *Behavior Genetics.* 35(3): 279-295.

65. Fire A, Xu S, Montgomery MK, Kostas S, Driver SE, Mello CC. (1998). Potent and specific genetic interference by a double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 391(6669): 806-811.

66. Forsyth DJ. (1968). The structure of the defence glands in the Dytiscidae, Noteridae, Haliplidae and Gyrinidae (Coleoptera). *Transactions of the Royal Entomological Society of London*. 120(6): 159-181.

67. Forsyth DJ. (1970). The structure of the defence glands of the Cicindelidae, Amphizoidae, and Hygrobiidae (Insecta: Coleoptera). *Journal of Zoology.* 160(1): 51-69.

68. Forsyth DJ. (1972). The structure of the pygidial defence glands of Carabidae (Coleoptera). *Transactions of the Zoological Society of London.* 32: 249-309.

69. Francke W and Dettner K. (2005). Chemical Signalling in Beetles. In: Schulz S. (eds.) *The Chemistry of Pheromones and Other Semiochemicals II*. *Topics in Current Chemistry*. Vol. 240. (pp. 85-166). Springer, Berlin, Heidelberg.

70. Fuller E, Elderd BD, and Dwyer, G. (2012). Pathogen Persistence in the Environment and Insect-Baculovirus Interactions: Disease-Density Thresholds, Epidemic Burnout and Insect Outbreaks. *The American Naturalist*. 179(3): E70-96.

71. Gamberale G and Tullberg BS. (1998). Aposematism and gregariousness: the combined effect of group size and coloration on signal repellence. *Proceedings of the Royal Society B: Biological Sciences.* 265(1399): 889–894.

72. Garry CE. (1993). Ground Beetles (Coleoptera: Carabidae) of Paleoenvironmental Significance of the Forest-Tundra and Open Woodland of Northern Manitoba, Canada. *The Coleopterists Bulletin* 47(1): 89-106.

73. Geiselhardt S, Jakobschy D, Ockenfels P, and Peschke K. (2008). A sex pheromone in the desert tenebrionid beetle *Parastizopus armaticeps*. *Journal of Chemical Ecol*ogy. 34(8): 1065–1071.

74. Geiselhardt SF, Geiselhardt S, and Peschke K. (2006). Chemical mimicry of cuticular hydrocarbons – how does *Eremostibes opacus* gain access to breeding burrows of its host *Parastizopus armaticeps* (Coleoptera, Tenebrionidae)? *Chemoecology*. 16(1): 59-68.

75. Geiselhardt SF, Peschke K, and Nagel P. (2007). A review of myrmecophily in ant nest beetles (Coleoptera: Carabidae: Paussinae): linking early observations with recent findings. *Naturwissenschaften*. 94(11): 871-894.

76. Gunawardena NE and Bandumathie MK. (1993). Defensive secretion of rice bug, *Leptocorisa oratorius* fabricus, (Hemiptera: Coreidae: A unique chemical combination and its toxic, repellent, and alarm properties. *Journal of Chemical Ecology.* 19(4): 851-861.

77. Hallem EA, Dahanukar A, and Carlson JR. (2006). Insect odor and taste receptors. *Annual Review of Entomol*ogy. 51: 113-135.

78. Hamilton WD. (1964a). The genetical evolution of social behavior. I. *Journal of Theoretical Biology.* 7(1): 1-16.

79. Hamilton WD. (1964b). The genetical evolution of social behavior. II. *Journal of Theoretical Biology.* 7(1): 17-52.

80. Hamilton WD. (1971). Geometry for the selfish herd *Journal of Theoretical Biology*. 31(2): 295-311.

81. Hamilton WD. (1972). Altruism and Related Phenomena, Mainly in Social Insects. *Annual Review of Ecology and Systematics.* 3: 193-232.

82. Hansson BS and Stensmyr MC. (2011). Evolution of Insect Olfaction. *Neuron*. 72(5): 698-711.

83. Holliday AE, Mattingly TM, and Holliday NJ. (2015). Defensive secretions of larvae of a carabid beetle. *Physiological Entomology* 40(2): 131-137.

84. Ioannou CC, Morell LJ, Ruxton GD, and Krause J. (2009). The effect of prey density on predators: conspicuousness and attack success are sensitive to spatial scale. *The American* Naturalist. 173(4): 499-506.

85. James A, Morison K, Todd S. (2012). A mathematical model of the defence mechanism of a bombardier beetle. *Journal of the Royal Society Interface*. 10(79): 20120801.

86. Jennings JH, Mazzi D, Ritchie MG, and Hoikkala A. (2011). Sexual and postmating reproductive isolation between allopatric *Drosophila montana* populations suggest speciation potential. *BMC Evolutionary Biology.* 11(68).

87. Jennings JH, Snook RR, and Hoikkala A. (2014). Reproductive isolation among allopatric *Drosophila montana* populations. *Evolution*. 68(11): 3095-3108.

88. Johnson CA, Vander Meer RK, Lavine B. (2001). Changes in the Cuticular Hydrocarbon Profile of the Slave-Maker Ant Queen, *Polyergus breviceps* Emery, After Killing a Formica Host Queen (Hymenoptera: Formicidae). *Journal of Chemical Ecology.* 27(9): 1787–1804.

89. Jones G and Teeling EC. (2006). The evolution of echolocation in bats. *Trends in Ecology & Evolution* 21(3): 149-156.

90. Juliano SA. (1984). Multiple feeding and aggression among larvae of *Brachinus lateralis* Dejean (Coleoptera: Carabidae). *The Coleopterists Bulletin.* 38(4): 358-360.

91. Kárpáti Z, Tasin M, Cardé RT, and Dekker T. (2013). Early quality assessment lessens pheromone specificity in a moth. *Proceedings of the National Academy of Sciences of the United States of America.* 110(18): 7377-7382.

92. Kistler KE, Vosshall LB, Matthews BJ. (2015). Genome Engineering with CRISPR-Cas9 in the Mosquito *Aedes aegypti*. *Cell Reports*. 11(1): 51-60.

93. Kleeberg I, Menzel F, and Foitzik S. (2017). The influence of slavemaking lifestyle, caste and sex on chemical profiles in *Temnothorax* ants: insights into the evolution of cuticular hydrocarbons. *Proceedings of the Royal Society B: Biological Sciences*. 284: 20162249.

94. Klun JA et al. (1975). Insect Sex Pheromones: Intraspecific Pheromonal Variability of *Ostrinia nubilalii* in North America and Europe. *Environmental Entomology*. 4(6). 891-894.

95. Klun JA, Chapman DL, Mattes KC, Wojtkowski PW, Beroza M, and Sonnet PE. (1973). Insect sex pheromones: minor amount of opposite geometrical isomer critical to attraction. *Science* 181(4100): 661–663.

96. Kocbansky J, Carde RT, Liebberr J, and Roelofs WL. (1975). Sex pheromones of the European corn borer, *Ostrinia nubialis* (Lepidoptera: Pyralidae), in New York. *Journal of Chemical Ecology*. 1(2): 225-231.

97. Landolt PJ. (1997). Sex Attractant and Aggregation Pheromones of Male Phytophagous Insects. *American Entomol*ogist 43(1): 12-22.

98. Lečić S, Ćurčić S, Vujisić L, Ćurčić B, Ćurčić N, Nikolić Z, Anđelković B, Milosavljević S, Tešević V, and Makarov S. (2014). Defensive secretions in three ground-beetle species (Insecta: Coleoptera: Carabidae). *Annales Zoologici Fennici.* 51(3): 285-300.

99. Li J, Lehmann S, Weißbecker B, Naharros IO, Schütz S, Joop G, and Wimmer EA. (2013). Odoriferous Defensive Stink Gland Transcriptome to Identify Novel Genes Necessary for Quinone Synthesis in the Red Flour Beetle, *Tribolium castaneum*. *PLOS Genetics*. 9(7): e1003596.

100. Lockwood JA and Story RN. (1987). Defensive secretion of the southern green stink bug (Hemiptera: Pentatomidae) as an alarm pheromone. *Annals of the Entomological Society of America.* 80(5): 686-691.

101. Löfqvist, J. (1976). Formic acid and saturated hydrocarbons as alarm pheromones for the ant *Formica rufa*. *Journal of Insect Physiology.* 22(10): 1331–1346.

102. Machado G, Bonato V, and Oliveria PS. (2002). Alarm communication: A new function for the scent-gland secretion in harvestmen (Arachnida: Opiliones). *Naturwissenschaften.* 89(8): 357-360.

103. Martin SJ, Vitikainen E, Helanterä H, and Drijfhout FP. (2008). Chemical basis of nest-mate discrimination in the ant *Formica exsecta*. *Proceedings of the Royal Society B: Biological Sciences*. 275(1640): 1271-1278.

104. Mast JD, De Moraes CM, Alborn HT, Lavis LD, and Stern DL. (2014). Evolved differences in larval social behavior mediated by novel pheromones. *Elife.* 3: e04205.

105. McDonald DB. (1989). Correlates of male mating success in a lekking bird with male-male cooperation. *Animal Behavior.* 37(6): 1007-1022.

106. Miller DR, Borden JH, and Lindgren SB. (1995). Verbenone: Dose-Dependent Interruption of Pheromone-Based Attraction of Three Sympatric Species of Pine Bark Beetles (Coleoptera: Scolytidae). *Environmental Entomology*. 24(3): 692-696.

107. Miller DR, Lindgren SB, and Borden JH. (2005). Dose-Dependent Pheromone Responses of Mountain Pine Beetle in Stands of Lodgepole Pine. *Environmental Entomology.* 34(5): 1019-1027.

108. Moore BP and Wallbank BE. (1968). Chemical composition of the defensive secretion in carabid beetles and its importance as a taxonomic character. *Systematic Entomology.* 37(5-6): 62-72.

109. Moore W, Song XB, and Di Giulio A. (2011). The larva of *Eustra* (Coleoptera: Paussinae, Ozaenini): A facultative associate of ants. *ZooKeys*. 90: 63-82.

110. Mori A, Grasso DA, Visicchio R, and Le Moli F. (2000a). Colony founding in *Polyergus rufescens*: the role of the Dufour's gland. *Insectes Sociaux*. 47(1): 7-10.

111. Mori A, Visicchio R, Sledge MF, Grasso DA, Le Moli F, Turillazzi S, Spencer S, and Jones GR. (2000b). Behavioural assays testing the appeasement allomone of

Polyergus rufescens queens during host-colony usurpation. *Ethology Ecology and Evolution.* 12(3): 315-322.

112. Nadeau JA, Petereit J, Tillett RL, Jung K, Fotoohi M, MacLean M, Young S, Schlauch K, Blomquist GJ, and Tittiger, C. (2017). Comparative transcriptomics of mountain pine beetle pheromone-biosynthetic tissues and functional analysis of CYP6DE3. *BMC Genomics*. 18:331.

113. Nanda P and Singh BN. (2012). Behavioural reproductive isolation and speciation in *Drosophila*. *Journal of Biosciences.* 37(2): 359-374.

114. Ohlen M, Herfurth AM, Kerbstadt, Wittstock U. (2016). Cyanide detoxification in an insect herbivore: Molecular identification of β-cyanoalanine synthases from Pieris rapae. *Insect Biochemistry and Molecular Biology*. 70: 99-110.

115. Okumura E and Yoshiga T. (2014). Host orientation using volatiles in the phoretic nematode *Caenorhabditis japonica*. *Journal of Experimental Biology* 217: 3197-3199.

116. Philip BA and Burgess EPJ. (2012). Observations on the ecology and behaviour of *Ctenognathus novaezelandiae* (Fairmaire) (Coleoptera: Carabidae). *New Zealand Entomologist*. 31(1): 35-39.

117. Raffa KF. (2011). Mixed messages across multiple trophic levels: the ecology of bark beetle chemical communication systems. *Chemoecology*. 11: 49–65.

118. Robertson JA and Moore W. (2016). Phylogeny of *Paussus* L. (Carabidae: Paussinae): unravelling morphological convergence associated with myrmecophilous life histories. *Systematic Entomology.* 42(1): 134-170.

119. Roelofs WL and Bjostad L. (1984). Biosynthesis of Lepidopteran pheromones. *Bioorganic chemistry*. 12(4). 279-298.

120. Roelofs WL and Rooney AP. (2003). Molecular genetics and evolution of pheromone biosynthesis in Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America.* 100(16): 9179-9184.

121. Roelofs WL, Du JW, Tang XH, Robbins PS, and Eckenrode CJ. (1985). Three European corn borer populations in New York based on sex pheromones and voltinism. *Journal of Chemical Ecology*. 11(7): 829-836.

122. Roelofs WL, Liu WT, Hao GX, Jiao HM, Rooney AP, and Linn CE. (2002). Evolution of moth sex pheromones via ancestral genes. *Proceedings of the National Academy of Sciences of the United States of America*. 99(21): 13621–13616.

123. Ruther J, Reinecke A, Tolasch T, and Hilker M. (2001). Make love not war: a common arthropod defence compound as sex pheromone in the forest cockchafer *Melolontha hippocastani. Oecologia.* 128(1): 44–47.

124. Salt G. (1928). Notes on the life history of *Pelecium sulcatum* Guerin. *Psyche*. 35: 131-134.

125. Sasahara K, Cody ML, Cohen D, Taylor CE. (2012). Structural Design Principles of Complex Bird Songs: A Network-Based Approach. *PLOS One*. 7(9). e44436.

126. Sasakawa K, Ikeda H, Sutou M, Dobata S, and Ito M. (2011). Parasitism of Adult *Poecilus versicolor* (Coleoptera: Carabidae) by Hymenopteran Larvae. The *Canadian Entomoogist.* 143(3): 211-223.

127. Saska P and Honek A. (2004). Development of the beetle parasitoids, *Brachinus explodens* and *B. crepitans* (Coleoptera: Carabidae). *Journal of Zoology.* 262(1): 29-36.

128. Schildknecht H. (1970). The defensive chemistry of land and water beetles. *Angewandte Chemie International Edition*. 9(1): 1–9.

129. Schildknecht HW, Maschwitz U, and Winkler H. (1968a). Über Arthropoden-Abwehrstoffe XXXII. Zur Evolution der Carabiden-Wehrdrüsensekrete. *Naturwissenschaften.* 55: 112–117.

130. Schildknecht HW, Winkler H, and Maschwitz U. (1968b). Über Arthropoden-Abwehrstoffe XXXI. Vergleichend chemische Untersuchungen der Inhaltsstoffe der Pygidialwehrblasen von Carabiden. *Zeitschrift für Naturforschung B.* 23: 637-644.

131. Steiner AM, Busching C, Vogel H, Wittstock U. (2018). Molecular identification and characterization of rhodaneses from the insect herbivore Pieris rapae. *Scientific Reports*. 8: 10819.

132. Stelicht M. (1973). Parasitic wasps attracted by the sex pheromone of the coccid host. *Entomophaga*. 18(4): 339-342.

133. Strand MR and Obrycki JJ. (1996). Host Specificity of Insect Parasitoids and Predators. *BioScience* 46(6): 422-429.

134. Sturgis SJ and Gordon DM. (2012). Nestmate recognition in ants (Hymenoptera: Formicidae): A review. *Myrmecological News*. 16: 101-110.

135. Symonds MRE and Elgar MA. (2003). The mode of pheromone evolution: evidence from bark beetles. *Proceedings of the Royal Society B: Biological Sciences.* 271(1541): 839-846.

136. Symonds MRE and Elgar MA. (2007). The evolution of pheromone diversity. *Trends in Ecology and Evolution.* 23(4): 220-228.

137. Talarico F, Bonacci T, Brandmayr P, Dalpozza R, De Nino A, Giglio A, Tagarelli A, and Brandmayr TZ. (2009). Avoiding ant detection in *Siagona europaea* Dejean 1826 (Coleoptera Carabidae): an evolutionary step towards true myrmecophily. *Ethology Ecology & Evolution.* 21(1): 45-61.

138. Teal PE & Tumlinson JH. (1988). Properties of cuticular oxidases used for sex pheromone biosynthesis by *Heliothis zea*. *Journal of Chemical Ecology.* 14(11): 2131-2145.

139. Teal PE and Tumlinson JH. (1986). Terminal steps in pheromone biosynthesis by *Heliothis virescens* and *H. zea*. *Journal of Chemical Ecology.* 12(2): 353-366.

140. Thomas CFG, Parkinson L, Griffiths GJK, Fernandez-Garcia A, and Marshall EJP. (2001). Aggregation and temporal stability of carabid beetle distributions in field and hedgerow habitats. *Journal of Applied Ecology.* 38(1): 100-116.

141. Tillman JA, Seybold SJ, Jurenka RA, Blomquist GJ. (1999). Insect pheromones—an overview of biosynthesis and endocrine regulation. *Insect Biochemistry and Molecular Biology.* 29(6): 481-514.

142. Trivers RL and Hare H. (1976). Haplodiploidy and the Evolution of the Social Insect. *Science*. 191(4224): 249-263.

143. Vander Meer RK and Wojcik DP. (1982). Chemical Mimicry in the Myrmecophilous Beetle *Myrmecaphodius excavaticollis*. *Science*. 218(4574): 806-808.

144. Visicchio R, Sledge MF, Mori A, Grasso DA, Le Moli F, Turillazzi S, Moneti G, Spencer S, and Jones GR. (1999). Dufour's gland contents of queens of the slave-making ant *Polyergus rufescens* and its host species *Formica cunicularia*. *Ethology Ecology & Evolution.* 12(1): 67-73.

145. Vogel H, Heidel AJ, Heckel DG, and Groot AT. (2010). Transcriptome analysis of the sex pheromone gland of the noctuid moth *Heliothis virescens*. *BMC Genomics*. 11(29).

146. Walgenbach CA, Phillips JK, Faustini DL, and Burkholder WE. (1982). Male-produced aggregation pheromone of the maize weevil, *Sitophilus zeamais*, and interspecific attraction between three *Sitophilus* species. *Journal of Chemical Ecology.* 9(7): 831-841.

147. Weber DC, Saska P, and Chaboo CS. (2008). Carabid Beetles (Coleoptera: Carabidae) as Parasitoids. In: Capinera JL (ed.) *Encyclopedia of Entomology*, Springer, Netherlands. 719–721.

148. Weber I. (1973). Tactile communication among free-range langurs. *American Journal of Physical Anthropology.* 32(2): 481-486.

149. Weiss I, Rössler T, Hofferberth J, Brummer M, Ruther J, and Stökl J. (2013). A nonspecific defensive compound evolves into a competition avoidance cue and a female sex pheromone. *Nature Communications.* 4:2767.

150. Wertheim B, van Baalen EJA, Dicke M, and Vet LEM. (2005). Pheromone-mediated aggregation in nonsocial arthropods: an evolutionary ecological perspective. *Annual Review of Entomology.* 50: 321–346.

151. Wicker-Thomas C. (2009). Evolution of insect pheromones and their role in reproductive isolation and speciation. *Annales de la Société entomologique de France.* 47(1-2): 55-62.

152. Will KW, Attygalle AB, and Herath K. (2000). New defensive chemical data for ground beetles (Coleoptera: Carabidae): interpretations in a phylogenetic framework. *Biological Journal of the Linnean Society.* 71(3): 459-481.

153. Wilson EO and Regnier FE. (1971). The Evolution of the Alarm-Defense System in the Formicine Ants. *The American Naturalist.* 105(943): 279-289.

154. Wölfling M and Rostás M. (2009). Parasitoids use chemical footprints to track down caterpillars. *Communicative & Integrative Biology.* 2(4): 353-355.

155. Xu T, Yasui H, Teale SA, Fujiwara-Tsujii N, Wickham JD, Fukaya M, Hansen L, Kiriyama S, Hao D, Nakano A, Zhang L, Watanabe T, Tokoro M, and Millar JG. (2017). Identification of a male-produced sex-aggregation pheromone for a highly invasive cerambycid beetle, *Aromia bungii*. *Scientific Reports*. 7: 7330.

156. Yew JY and Chung H. (2015). Insect pheromones: An overview of function, form, and discovery. *Progress in Lipid Research.* 59: 88-105.

157. Young S, Watt PJ, Grover JP, and Thomas D. (1994). The Unselfish Swarm? *Journal of Animal Ecology.* 63(3): 611-618.

158. Zhang G, Chen J, Yu H, Tian X, Wu J. (2018). Molecular and Functional Characterization of pheromone binding protein 1 from the Oriental Fruit Moth, Grapholita molesta (Busck). *Scientific Reports.* 8: 2276

159. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, Leonhardt H, Heyn H, Hellmann I, Enard W. (2017). Comparative Analysis of Single-Cell RNA Sequencing Methods. *Molecular Cell*. 65(4): 631-643.

FIGURES AND TABLES

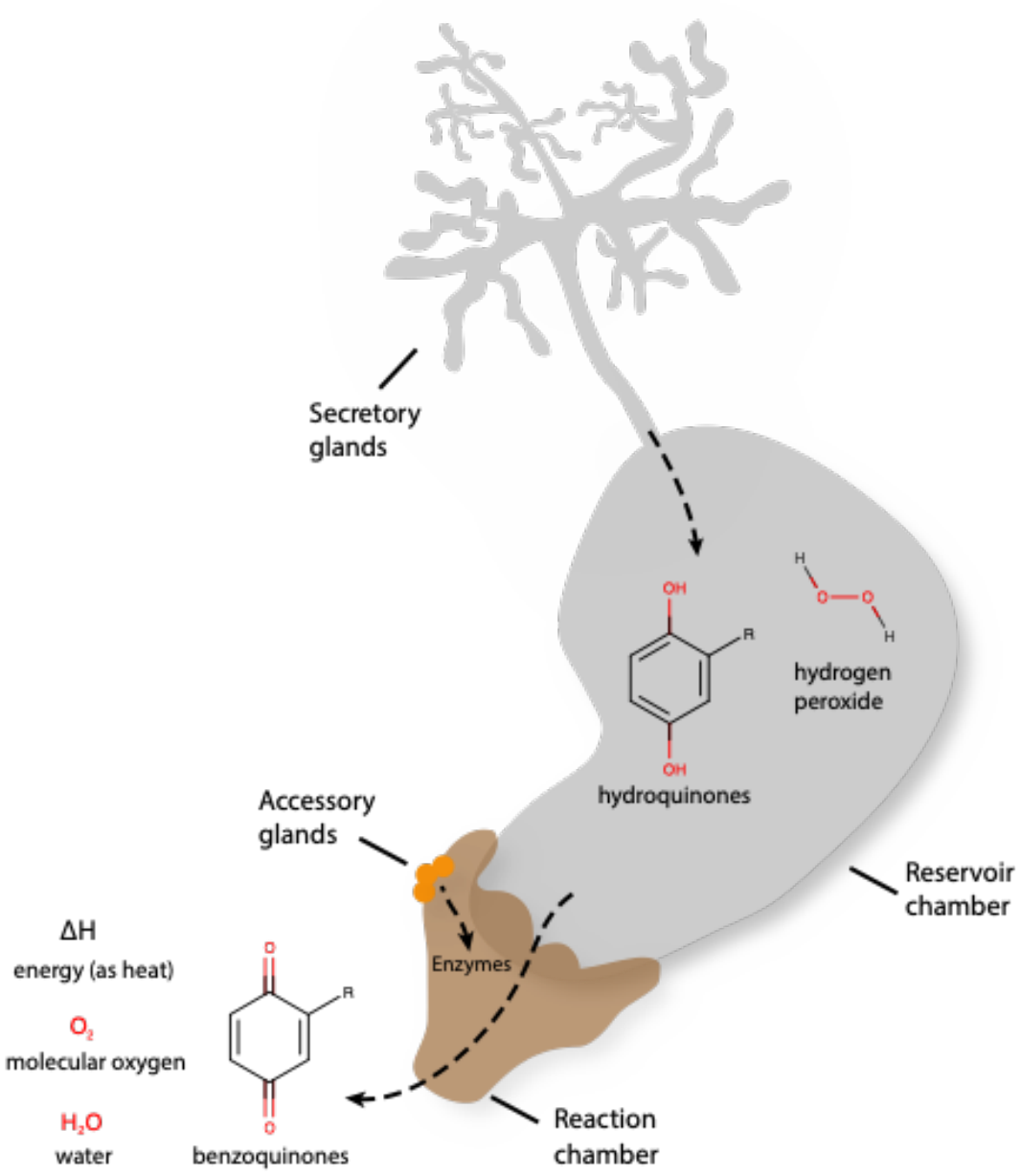


Figure 1. A diagram showing a pygidial gland reservoir (res) and reaction chamber (rc) of a bombardier beetle, adapted from (James et al. 2012). The pygidial gland reservoir contains a mixture of hydrogen peroxide and hydroquinones, which upon reservoir contraction, are ejected into a sclerotized reaction chamber through a one-way value

(val). In the reaction chamber are enzymes produced by accessory glands (ag) which catalyze the degradation of hydrogen peroxide and convert hydroquinones to benzoquinones. The enthalpy of this process is great enough to cause a pressure gradient which expels the benzoquinones out of the abdomen at temperatures reaching up to 100°C.

Figure 2. Hypothetical depiction of chemically-mediated tritrophic interactions between carabid beetles, a wolf spider, and a parasitoid. (A) Carabid spraying the wolf spider, a predator, with its pygidial gland allomones; (B) Pygidial gland allomones acting as alarm pheromones for conspecifics; (C) A carabid beetle releasing an aggregation pheromone, which is detected by a conspecific at the other side of the rock; (D) A parasitoid wasp eavesdropping on volatile cues (kairomones) released by aggregating carabid beetles.

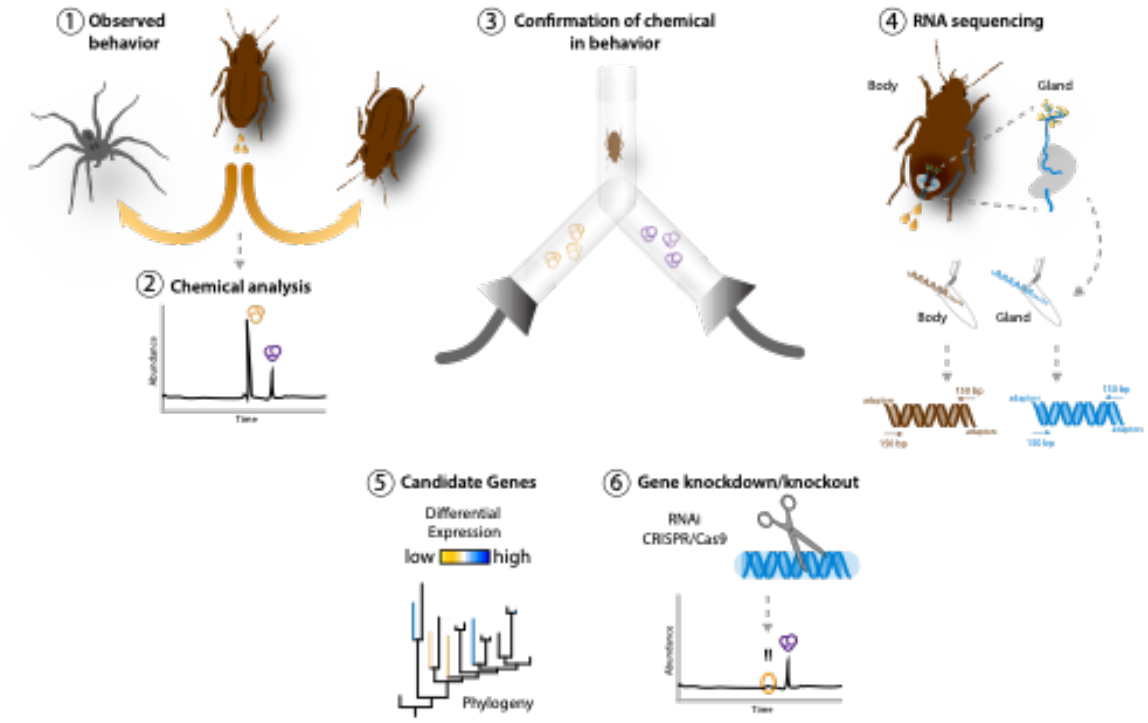
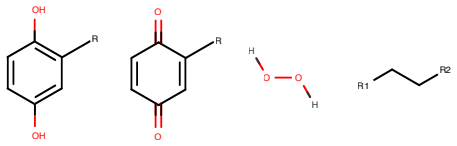
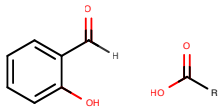
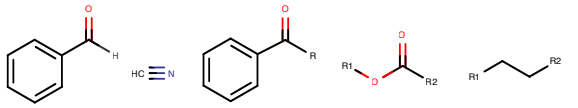
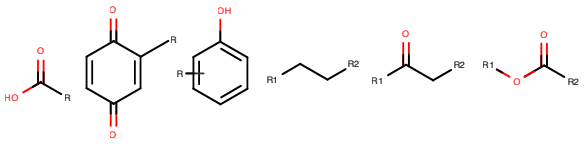
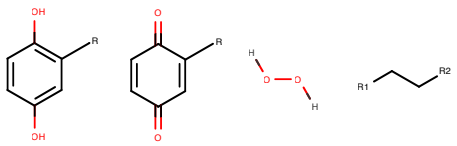
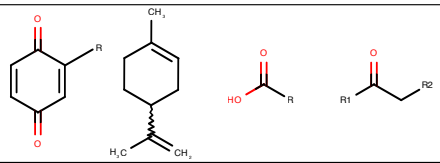


Figure 3. Workflow showing how methods from ethology, chemistry, and molecular biology can be integrated toward the goal of elucidating the evolution of chemically-mediated insect-insect interactions. (A) Behavioral observations are made of a carabid beetle spraying pygidial gland contents, which are detected by a spider (allomone) and a conspecific (alarm pheromone). (B) Interested in what chemicals are mediating this interaction, GC-MS analyses are first conducted on gland extracts. (C) Behavioral assays are then conducted using individual gland constituents or blends to characterize their role in behavior (if any). (D) RNA is extracted from glandular and non-glandular tissue and is sent for transcriptome sequencing. (E) Differentially expressed candidate genes that may be involved in the biosynthesis of behaviorally active compounds are identified. Phylogenetic trees of gene families may be reconstructed to infer the evolution of semiochemical production within a clade. (F) Gene knockdown or gene knockout is used

to confirm the role of candidate gene products in the biosynthesis of the semiochemical of interest. Subsequent analyses could be conducted as dictated by research interests.

REPRESENTATIVE PYGIDIAL GLAND ALLOMONES OF SELECT CARABIDAE SUBFAMILIES

Subfamily	Chemicals produced	References	Chemical structures
Brachininae	Quinones, hydrogen peroxide, hydrocarbons	Schildknecht et al. 1964, Moore & Wallbank 1968, Eisner et al. 1977, Kanehisa & Murase 1977, Eisner et al. 2001	
Carabinae	Salicylaldehyde, aliphatic organic acids	Eisner et al. 1963, Schildknecht et al. 1964, McCullough 1966, Moore & Wallbank 1968, Schildknecht 1970	
Cicindelinae	Benzaldehyde, hydrogen cyanide, aromatic acids, aliphatic esters, hydrocarbons	Blum et al. 1981, Pearson et al. 1998, Kelley & Schilling 1998, Schultz & Puchalski 2001	
Harpalinae	Aliphatic organic acids, quinones, phenolics, hydrocarbons, aliphatic ketones, aliphatic esters	Moore & Wallbank 1968, Kanehisa & Murase 1977, Kanehisa & Kawazu 1985, Attygalle et al. 1992, Will et al. 2000, Holliday et al. 2012	
Paussinae	Quinones, hydrogen peroxide, hydrocarbons	Moore & Wallbank 1968, Eisner et al. 1977, Roach et al. 1979, Aneshansley et al. 1983, Eisner et al. 1989	
Scaratinae	Quinones, terpenes, aliphatic organic acids, aliphatic ketones	Moore & Wallbank 1968, Kanehisa & Murase 1977, Moore & Brown 1979, Attygalle et al. 2009	

Additional Information:
1. Compounds shown for each subfamily do not represent an exhaustive list of compounds produced. Those represented here are amongst the most abundant or are biochemically unique within Carabidae.
2. "R", "R1", and "R2" represent any functional group of a given length. Given the diversity of carabid compounds, the pygidial gland constituents can be best represented in terms of chemical classes.
3. An R-group intersecting a bond within a ring indicates that the R-group may bind to multiple different carbons (i.e. there is more than one compound with that general structure).

Table 1. A table demonstrating common pygidial gland chemical classes produced by six major Carabidae subfamilies. Chemical structures are also presented for each compound in order of their listing.

CHAPTER 2

Pygidial glands of *Harpalus pensylvanicus* (Coleoptera: Carabidae) contain resilin-rich structures.

This work has been previously published and has been reproduced with permission from Elsevier (https://doi.org/10.1016/j.asd.2018.12.004).

ABSTRACT

The pygidial gland system is a key innovation in adephagan beetles, producing, storing, and spraying defensive chemical compounds. As the source of defensive chemical production and storage, the pygidial gland system experiences severe chemical stress which challenges the integrity of the entire gland system. Here, we utilize autofluorescence-based confocal laser scanning microscopy to examine the morphology of pygidial gland secretory lobes and collecting ductules in a common Pennsylvanian harpaline species, *Harpalus pensylvanicus*. The glandular units are composed of type-III exocrine cells which empty into resilin-rich ductules, which themselves lead into a larger resilin-rich collecting duct, and ultimately the pygidial reservoir pump. We also utilize histological staining with toluidine blue and brightfield imaging to provide additional support for the presence of resilin in the collecting duct, as toluidine blue has been shown to stain resilin without metachromasia. We hypothesize that the high resilin content of the collecting ducts might be a widespread key evolutionary adaptation to prevent damage caused by physical and chemical stress generated in pump-containing insect exocrine gland systems.

INTRODUCTION

The family Carabidae (Coleoptera: Adephaga) represents a hyperdiverse lineage of insects at almost 40,000 described species; nearly 10% of all described Coleopterans (Beutel and Leschen, 2005). A unique synapomorphy of adephagan beetles is the paired pygidial gland system, an invagination of the last tergal epidermis, which is used in most taxa for defense against predators (Forsyth, 1968, 1970, 1972; Dettner, 1985; Beutel and Haas, 1996). Located in the abdomen, the pygidial glands consist of defensive chemical-producing secretory lobes, which are connected to a musculated reservoir chamber via a long collecting duct (Will et al., 2000; Attygalle et al., 2007). In most Carabidae, this reservoir opens to the apex of the abdomen via an efferent duct.

The gross pygidial gland morphology across taxa examined thus far reveals an astounding array of diversity in almost every aspect of the system. In some species, secretory lobes are composed of fewer than 20 elongated structures, while those in other species are composed of well over 200 small glandular spheres (Forsyth, 1972; Kanehisa and Murase, 1977). The reservoir chambers are similar across most taxa, almost all being ovular to kidney-shaped in appearance. Some formic acid-producing taxa have an additional dorsal lobe on the reservoir that is thought to hold the surfactant lipid phase of the defensive spray (Will et al., 2000). The efferent duct is absent in the Brachininae and Paussinae, which instead have evolved a highly sclerotized reaction chamber with its own glandular system for producing enzymes that catalyze a reaction between hydrogen peroxide and hydroquinones to produce benzoquinones, molecular oxygen, and water (Di Giulio et al., 2015; Eisner et al., 2000; Schildknecht et al., 1968a,b; Schildknecht and Koob, 1969; Balestrazzi et al., 1985).

Most carabid beetles can spray their pygidial gland contents at a distance multiple times their body length (Forsyth, 1972). Defensive spray systems are known from some other insect lineages, such as the prothoracic gland in Phasmatidae (Phasmatodea) (Eisner, 1965) and the abdominal defensive glands in Tenebrionidae (Coleoptera) (Kendall, 1974). These exocrine gland systems always rely on a reservoir pump component, and thus may experience physical stress from the pressure they generate. So far, no studies have suggested how insects are able to prevent damage to gland components caused by potential stress generated from reservoir pump contractions or the defensive chemicals they contain.

In the present study, we aim to describe the gross morphology of the pygidial gland secretory lobes of a common Pennsylvanian carabid, *Harpalus pensylvanicus*, with confocal laser scanning microscopy (CLSM) using autofluorescence, histological staining, and brightfield imaging (Figs. 1-4).

MATERIALS AND METHODS

Insect Collections

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 7 October 2017, and from 01 September 2018 to 09 October 2018. Individuals were transported to campus post-collection where they were placed in terrariums with soil from the field site and coconut fiber substrate (ZooMed Laboratories Inc). 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 deceased beetles were removed from the terrariums as they were discovered.

Dissections

Harpalus pensylvanicus dissections took place at the Pennsylvania State University (State College, PA, USA) using Olympus SZX16 and SZ61 stereomicroscopes. To aid in stabilizing the specimens during dissection as needed, individuals were pinned through the prothorax into silicon embedded Petri dishes (Sylgard 184 Silicone™), ventral side up. Otherwise, they were dissected in a watch glass without being pinned. 0.1M sodium phosphate (Fisher Chemical & EMD Chemicals) or sodium cacodylate buffer (Electron Microscopy Sciences) was poured over individuals until they were fully submerged. Incisions were made from the last abdominal sternite through the metasternum, and through the pleural membranes of the abdomen. The sterna were pinned back or removed entirely, and tissue was removed until the pygidial gland system became visible. One or both of the gland systems were removed from each dissected beetle, either as fully intact structures or as individual portions. Intact gland systems were subsequently cut into components, namely the secretory lobes, the collecting duct, the reservoir, and the efferent duct. Secretory lobes were further partitioned into smaller sections to aid in visualization using CLSM and brightfield microscopy.

CLSM Fixation

A fixative consisting of 1 mL 25% glutaraldehyde (Electron Microscopy Sciences), 5 mL 0.2M sodium phosphate buffer, and 0.5 g sucrose (Fisher Chemical) was prepared and

diluted to 10 mL using Nanopure water. Gluteraldehyde has been shown to increase the autofluorescence of tissues in a number of different systems (Collins and Goldsmith, 1981; Fester et al. 2008). Without the addition of glutaraldehyde, the secretory lobe tissues would likely not have fluoresced as intensely under laser excitation. Gland components that were determined to be of suitable quality were placed into fractions of this fixative for 72 hours. Only the secretory lobes and attached collecting ducts were selected for this study and were transferred from the fixative to glycerol droplets (Electron Microscopy Sciences) on a 1.5 glass coverslip (VWR) with Blu-tack (Bostic™) placed in each corner. A second 1.5 glass coverslip was placed on the Blu-tack and was pressed down until the glycerol droplet adhered to the surface of both, suspending the gland components (Miko and Deans, 2013).

CLSM Imaging

Glass coverslips containing the fixed secretory lobes were placed into an Olympus FV10i-LIV Confocal Laser Scanning Microscope at the Microscopy and Cytometry Facility at the Pennsylvania State University (State College, PA, USA). Samples were scanned using three excitation wavelengths: 405 nm (violet), 473 nm (blue), and 559 nm (green). Narrow emission wavelengths were 420-520 nm, 490-540 nm, and 570-620 nm respectively. The use of the 405 nm (violet) laser allowed us to visualize resilin present in the dissected tissues while the blue and green lasers show the presence of epithelial tissue (Michels and Gorb, 2012). Scanned image files were rendered and processed using FIJI (Fiji Is Just Image-J), (Schneider et al., 2012). Pseudocolors “Blue”, “Green”, and “Red”

were applied to channels 1 (420-520 nm), 2 (490-540 nm), and 3 (570-620 nm) respectively.

Histological Staining & Brightfield Imaging

A freshly dissected, unstained section of secretory lobe was mounted in a droplet of 0.1M sodium phosphate buffer between two 1.5 glass coverslips as described above. The gland section was immediately imaged at 10x magnification using an Olympus BX61 microscope equipped with a Hamamatsu Orca-ER C4742-80 digital camera at the Microscopy and Cytometry Facility at the Pennsylvania State University (State College, PA, USA). Post-imaging, the secretory lobe section was carefully removed from the coverslip and stained in a toluidine blue solution as a secondary test for the presence of resilin (Andersen and Weis-Fogh, 1964).

The 0.005% toluidine blue stock solution was prepared by dissolving 1 mg of toluidine blue dye powder (Electron Microscopy Sciences) in 20mL of Nanopure water. The secretory lobe section was transferred to a 2 mL aliquot of this stain and left at room temperature for 18 hours to allow for full uptake of the dye into the lumen and glandular spheres. Stained sections were then washed in ~20 mL of 0.1M sodium phosphate buffer for 5 hours to remove excess dye from the duct lumina. Stained glands were mounted in a droplet of 0.1M sodium phosphate buffer between two 1.5 glass coverslips as described above. Prior to imaging, the droplet of buffer containing the stained gland tissue was inspected against a white background to ensure that no excess dye was present in the sample post-wash. The gland section was then imaged at 10x magnification using the same Olympus BX61 microscope. We observed that although the original description of

this method by Andersen and Weis-Fogh suggested a 40-hour stain time, we were able to sufficiently stain our tissue using a simplified protocol (Andersen and Weis-Fogh, 1964).

RESULTS

The secretory lobes are comprised of a system of glandular spheres, each of which articulate to one ductule (sph, duc: Figs. 1, 2). These ductules subsequently lead into larger ducts, which all ultimately empty into a primary collecting duct that itself empties into a musculated reservoir chamber (cd: Fig. 1). Although the total number of secretory lobe glandular spheres varies by individual, most contain between 50 and 100 per gland system, or 100 to 200 per individual. The glandular spheres themselves are on average ~120 μm in diameter and autofluoresce when excited with blue and green lasers, shown here in green and red pseudocolor respectively (Fig. 1, 2). The orange colorization of parts of the images are due to colocalization of the autofluorescence shown in green and red pseudocolors.

Except for a thin, autofluorescing external layer shown in green pseudocolor (epi: Fig. 3), the collecting duct and ductules possess a strong autofluorescence when excited with the violet laser, shown in blue pseudocolor (Figs. 1-3). The lumina of the duct and ductules are surrounded by a 3–6 μm thick, blue-fluorescing tissue (lum: Fig. 1-3) that is connected to the duct wall via narrow lamellae arranged into a spiral, the cuticular crista (ccr: Fig. 3). Histological staining with toluidine blue reveals the lumen of the collecting duct to stain without metachromasia, indicating the presence of resilin (Andersen and Weis-Fogh, 1964) (Fig. 4a,b). We suggest that the impermeability of the collecting duct lumen to fluids (e.g. defensive chemicals) is the primary reason for a lack of stain uptake

into the rest of the duct tissue wall. The outermost layer of the duct does not appear to be comprised of resilin as is suggested by the lack of staining on the duct surface and by the lack of excitation at the surface by the CLSM violet laser. Rather, it is composed of an epithelial layer which itself is devoid of resilin (Figs. 2,3).

DISCUSSION

Carabid beetle pygidial glands may be considered a key innovation that has allowed one of the animal kingdom's most specious families to defend themselves against predators. While much work has been done to describe the gross and fine morphology of these glands in many subfamilies, there has been relatively little work done on their tissue composition (Forsyth, 1968, 1970, 1972; Di Giulio, 2015). This information is crucial to understanding not only how these structures evolved, either de novo or derived from an ancestral character state, but also how they withstand the stress of containing and ejecting cytotoxic chemicals.

Based on the strong autofluorescence under violet laser excitation (405 nm, shown in blue pseudocolor), we have determined that the cuticle of the collecting duct and ductules of the pygidial glands of *Harpalus pensylvanicus* is resilin-rich. This is supported by a histological assay using toluidine blue, which stains resilin without metachromasia as described by Andersen and Weis-Fogh (Weis-Fogh, 1960, Andersen and Weis-Fogh, 1964). Metachromasia refers to the phenomenon of a sample staining a color other than that of the stain used. Based on the lack of autofluorescence from violet laser excitation using CLSM, the glandular spheres of the secretory lobes do not contain resilin other than the ductules that they empty into. Resilin is among the most efficient

elastomeric proteins known (Elvin et al., 2005; Chen and Thouas, 2014; Jensen and Weis-Fogh, 1962). It is also stress-resistant and generally found in the conjunctiva to allow for movement of sclerites (Michels et al., 2016; Qin et al., 2012). It is also present in more specialized structures, such as the energy-storing pads in the hindlegs of Siphonaptera and Cercopidae, which allow them to jump multiple times their body length (Burrows et al., 2008; Burrows, 2009; Bennet-Clark and Lucey, 1967). It has even been found in non-insect animals, such as in the mandibles of the copepod *Centropages hamatus* (Michels, et al. 2012).

Here, we hypothesize that the presence of resilin in pygidial gland collecting ducts and ductules serves to manage the physical and/or chemical stress generated by the pygidial gland system. Were the collecting duct too rigid, any back pressure could damage delicate upstream structures, such as secretory lobe cells. If it were flexible, yet delicate (e.g. a single-layer epithelium), it may become damaged by back pressure itself. Our hypothesis is also supported by the fact that resilin provides flexibility for the spermathecal duct of some Dipteran species to prevent possible pressure damage caused by the contraction of the muscular spermathecal pump (Ilango, 2005; Clements and Potter, 1967). Since resilin is also highly resistant to chemical degradation, its presence in the collecting duct lumen may also facilitate the transport of cytotoxic and corrosive defensive chemicals from the secretory lobes to the reservoir (Chen and Thouas, 2014).

FUTURE DIRECTIONS

Two especially interesting groups for further CLSM study would be the bombardier beetles of the subfamilies Brachininae and Paussinae. Not only do their pygidial glands

have to cope with the production, transport, and storage of cytotoxic hydroquinones and hydrogen peroxide, but they also have to withstand the heat and pressure produced by the enzyme-catalyzed reaction between the two in their reaction chambers (Schildknecht et al., 1968a,b; Moore and Wallbank, 1968; Eisner et al., 1977; Will et al., 2000; Lečić et al., 2014). Given that some of these reactions have been recorded to occasionally reach 100°C, an understanding of their pygidial gland morphology and tissue composition is necessary to understanding the evolution of the bombardier beetle "blasting" ability (Aneshansley et al., 1969; Eisner et al., 1977; Dean et al., 1990; Beutel and Lawrence, 2005). One such morphological study has been carried out in *Brachinus elongatulus* and could be complimented by further studying these structures' compositions using techniques such as CLSM (Di Giulio et al., 2015).

The use of CLSM to study insect glands and contractile organs should not be limited to Carabidae. Just as resilin-rich pads used for locomotion have independently evolved in the Siphonaptera and Cercopidae, resilin-rich ductules may have evolved in other insect lineages with pump-driven defensive gland systems. Tenebrionidae and Phasmatodea, for example, have analogous defensive gland systems in which glands are interconnected by narrow ductules (Tschinkel, 1975; Tschinkel and Doyen, 1980; Abhitha et al., 2010; Kendall, 1974; Eisner, 1965). It is likely that these ductules also contain resilin, a hypothesis that is testable via the methods described in this paper and others (Weis-Fogh, 1960; Andersen and Weis-Fogh, 1964; Michels and Gorb, 2012).

Not only would further study into the tissue composition of arthropod exocrine glands shed new light on their evolution and function, but it would also give us a better understanding of how important tissue composition is to proper gland function, especially

under extreme physical and chemical stress. In the future, we plan to conduct CLSM and SBF-SEM-based morphological studies on the pygidial gland systems of other Carabidae taxa to better understand the tissue composition and fine morphology of these defensive structures across the family.

ACKNOWLEDGEMENTS

We thank Sulav Paudel, Asifa Hameed, Po-An Lin, Ching-Wen Tan, and Maria Perezsandi for thoughtful discussion and assistance with dissections. We also thank Andy Deans for constructive feedback, as well as Jonah Ulmer, Chloe Drummond, and Kylie Bocklund for thoughtful discussions on staining. Finally, we thank Missy Hazen and John Cantolina for their help with CLSM and Olympus BX61 imaging (Penn State Microscopy and Cytometry Facility - University Park, PA). This work was supported by the National Science Foundation (DEB 1556931/1762760 to Tanya Renner). The findings and conclusions of this work do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Abhitha, P., Vinod, K. and Sabu, T.K., 2010. Defensive Glands in the Adult and Larval Stages of the Darkling Beetle, Luprops tristis. J. Insect Sci. 10: 7.
2. Andersen, S.O. and Weis-Fogh, T., 1964. Resilin. A rubberlike protein in Arthropod Cuticle. Advances in Insect Physiology. 2: 1-65.
3. Aneshansley, D.J., Eisner, T., Widom, J.M. and Widom, B., 1969. Biochemistry at 100 °C: Explosive Secretory Discharge of Bombardier Beetles (Brachinus). Science. 165(3888): 61-63.
4. Attygalle, A.B., Wu, X. and Will, K.W., 2007. Biosynthesis of tiglic, ethacrylic, and 2-methylbutyric acids in a carabid beetle, Pterostichus (Hypherpes) californicus. J. Chem. Ecol. 33(5): 963–970.
5. Balestrazzi, E., Dazzini, M.V., de Bernardi, M., Vidari, G., Vita-Finzi, P. and Mellerio, G., 1985. Morphological and chemical studies on the pygidial defence glands of some Carabidae (Coleoptera). Naturwissenschaften. 72(9): 482–484.
6. Bennet-Clark, H.C., and Lucey, E.C.A., 1967. The jump of the flea: A study of the energetics and a model of the mechanism. J. Exp. Biol. 47(1): 59-76.
7. Beutel, R.G. and Haas, A., 1996. Phylogenetic analysis of larval and adult characters of Adephaga (Coleoptera) using cladistics computer programs. Insect Syst. Evol. 27(2): 197-205.
8. Beutel, R.G. and Lawrence, J.F., 2005. 4. Coleoptera, Morphology. In: Beutel, R.G. and Leschen, R.A.B. (Eds), Handbook of Zoology. Volume IV Arthropoda: Insecta. Part 38. Coleoptera. Volume 1: Morphology and Systematics (Archostemata,

Adephaga, Myxophaga, Polyphaga partim). Walter de Gruyter, Berlin, New York, pp. 23-27.

9. Burrows, M., Shaw, S.R., Sutton, G.P., 2008. Resilin and chitinous cuticle form a composite structure for energy storage in jumping by froghopper insects. BMC Biol. 6(41).
10. Burrows, M., 2009. How fleas jump. J. Exp. Biol. 212(18): 2881–2883.
11. Chen, Q. and Thouas, G., 2014. Chapter 10: Bioresorbable Polymers. In: Biomaterials: A Basic Introduction. CRC Press, Taylor Francis Group, Florida, pp. 360-361.
12. Clements, A. and Potter, S.A., 1967. The fine structure of the spermathecae and their ducts in the mosquito Aedes aegypti. J. Insect Physiol. 13(12): 1825-1828, IN3-IN12, 1829-1836.
13. Collins, J.S. and Goldsmith, T.H., 1981. Spectral properties of fluorescence induced by glutaraldehyde fixation. J. Histochem. Cytochem. 29(3): 411-444.
14. Dean, J., Aneshansley, D.J., Edgerton, H.E. and Eisner, T., 1990. Defensive spray of the bombardier beetle: a biological pulse jet. Science. 248(4960): 1219-1221.
15. Dettner, K., 1985. Ecological and phylogenetic significance of defensive compounds from pygidial glands of Hydradephaga (Coleoptera). Proc. Natl. Acad. Sci. Philadelphia. 137: 156-171.
16. Di Giulio, A., Muzzi, M., Romani, R., 2015. Functional anatomy of the explosive defensive system of bombardier beetles (Coleoptera, Carabidae, Brachininae). Arthropod Struc. Dev. 44(5): 468-490.
17. Eisner, T., 1965. Defensive spray of a phasmid insect. Science. 148(3672): 966–968.

18. Eisner, T., Aneshansley, D.J., Eisner, M., Attygalle, A.B., Alsop, D.W. and Meinwald, J., 2000. Spray mechanism of the most primitive bombardier beetle (Metrius contractus). J. Exp. Biol. 203: 1265-1275.
19. Eisner, T., Jones, T.H., Aneshansley, D.J., Tschinkel, W.R., Silberglied, R.E. and Meinwald, J., 1977. Chemistry of Defensive Secretions of Bombardier Beetles (Brachinini, Metriini, Oxaenini, Paussini). J. Insect. Physiol. 23: 1383-1386.
20. Elvin, C.M., Carr, A.G., Huson, M.G., Maxwell, J.M., Pearson, R.D., Vuocolo, T., Liyou, N.E., Wong, D.C., Merritt, D.J. and Dixon, N.E., 2005. Synthesis and properties of crosslinked recombinant pro-resilin. Nature. 437(7061): 999–1002.
21. Fester, T., Bergs, R.H., Taylor, C.G., 2008. An easy method using glutaraldehyde-introduced fluorescence for the microscopic analysis of plant biotrophic interactions. J. Microsc. 231(2): 342-348.
22. Forsyth, D.J., 1968. The structure of the defence glands in the Dytiscidae, Noteridae, Haliplidae and Gyrinidae (Coleoptera). Trans. R. Ent. Soc. Lond. 120: 158-181.
23. Forsyth, D.J., 1970. The structure of the defence glands of the Cicindelidae, Amphizoidae, and Hygrobiidae (Insecta: CoIeoptera). J. Zool. Lond. 160: 51-69.
24. Forsyth, D.J., 1972. The structure of the pygidial defence glands of Carabidae (Coleoptera). Trans. Zool. Soc. Lond. 32: 249-309.
25. Ilango, K., 2005. Structure and function of the spermathecal complex in the phlebotomine sandfly Phlebotomus papatasi Scopoli (Diptera: Psychodidae): I. ultrastructure and histology. J. Biosci. 30(5): 711–731.

26. Jensen, M., and Weis-Fogh, T., 1962. Biology and physics of locust flight. V. Strength and elasticity of locust cuticle. Philos. Trans. Royal Soc. B. 245(721): 137-169.

27. Kanehisa, K. and Murase, M., 1977. Comparative study of the pygidial defensive systems of carabid beetles. Appl. Entomol. Zool. 12(3): 225–235.

28. Kendall, D., 1974. The structure of defence glands in some Tenebrionidae and Nilionidae (Coleoptera). Ecol. Entomol. 125(4): 437–487.

29. Lečić, S., Ćurčić, S., Vujisić, L., Ćurčić, B., Ćurčić, N., Nikolić, Z., Anđelković, B., Milosavljević, S., Tešević, V. and Makarov, S., 2014. Defensive secretions in three ground-beetle species (Insecta: Coleoptera: Carabidae). Ann. Zool. Fennici. 51: 285-300.

30. Michels, J., Appel, E. and Gorb, S.N., 2016. Functional diversity of resilin in Arthropoda. Beilstein J. Nanotechnol. 7:1241-1259.

31. Michels, J. and Gorb, S.N., 2012. Detailed three-dimensional visualization of resilin in the exoskeleton of arthropods using confocal laser scanning microscopy. J. Microsc. 245(1): 1–16.

32. Michels, J., Vogt, J., Gorb, S.N., 2012. Tools for crushing diatoms – opal teeth in copepods feature a rubber-like bearing composed of resilin. Sci. Rep. 2: 465.

33. Mikó, I. and Deans, A., 2013. What is fluorescing? Hamuli. 4: 19–22.

34. Moore, B.P. and Wallbank, B.E., 1968. Chemical composition of the defensive secretion in carabid beetles and its importance as a taxonomic character. Syst. Entomol. 37(5-6): 62-72.

35. Qin, G., Hu, X., Cebe, P., Kaplan, D.L., 2012. Mechanism of resilin elasticity. Nature Communications 3: 1003.

36. Schildknecht, H.W. and Koob, K., 1969. Zur Explosionschemie der Bombardierkäfer. Naturwissenschaften. 56: 328.

37. Schildknecht, H.W., Maschwitz, U. and Winkler, H., 1968a. Zur Evolution der Carabiden-Wehrdrüsensekrete. Über Arthropoden-Abwehrstoffe XXXII. Naturwissenschaften. 55: 112–117.

38. Schildknecht, H.W., Winkler, H. and Maschwitz, U., 1968b. Vergleichend chemische Untersuchungen der Inhaltsstoffe der Pygidialwehrblasen von Carabiden. Über Arthropoden-Abwehrstoffe XXXI. Zeitschrift für Naturforschung B. 23: 637.

39. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W., 2012. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 9(7): 671–675.

40. Tschinkel, W.R., 1975. A comparative study of the chemical defensive system of tenebrionid beetles III. Morphology of the glands. J. Morphol. 145(3): 355–370.

41. Tschinkel, W.R. and Doyen, J.T., 1980. Comparative anatomy of the defensive glands, ovipositors and female genital tubes of tenebrionid beetles (Coleoptera). Int. J. Insect Morphol. Embryol. 9(5-6): 321–368.

42. Weis-Fogh, T., 1960. A Rubber-Like Protein in Insect Cuticle. J. Exp. Biol. 37: 889-907.

43. Will, K.W., Attygalle, A.B. and Herath, K., 2000. New defensive chemical data for ground beetles (Coleoptera: Carabidae): interpretations in a phylogenetic framework. Biol. J. Linn. Soc. 71: 459-481.

FIGURES AND TABLES

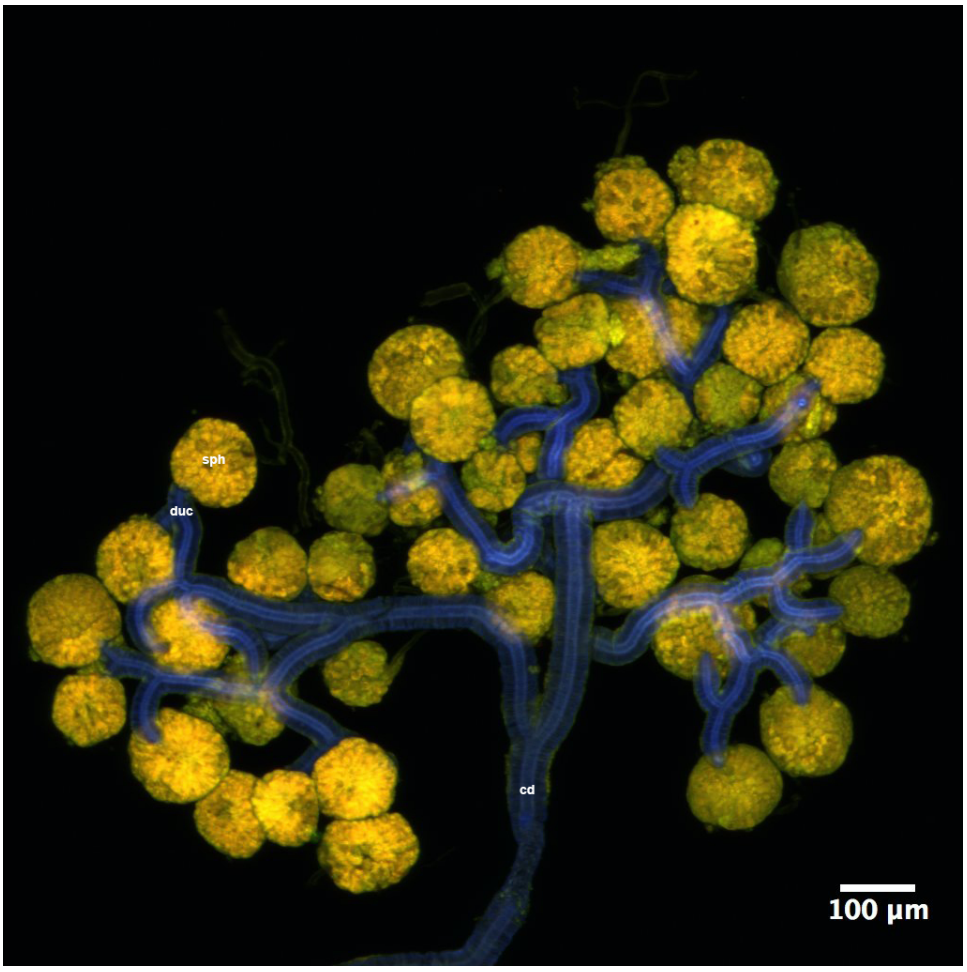


Figure 1. Secretory lobe section of *Harpalus pensylvanicus* viewed via CLSM. CLSM volume rendered micrograph showing the distal portion of the pygidial gland of *Harpalus pensylvanicus*. Glandular spheres (sph) show red-green autofluorescence while the collecting duct (cd) and ductules (duc) connecting the collecting duct with the glandular spheres autofluoresce in the blue interval showing the typical emission wavelength for resilin in response to the 405 nm (violet) excitation laser. Resilin-rich 3-6 µm thick tissue surrounds the lumen of the ductules and the collecting duct and is connected to the duct/ductule surface by lamellar structures, the cuticular crista.

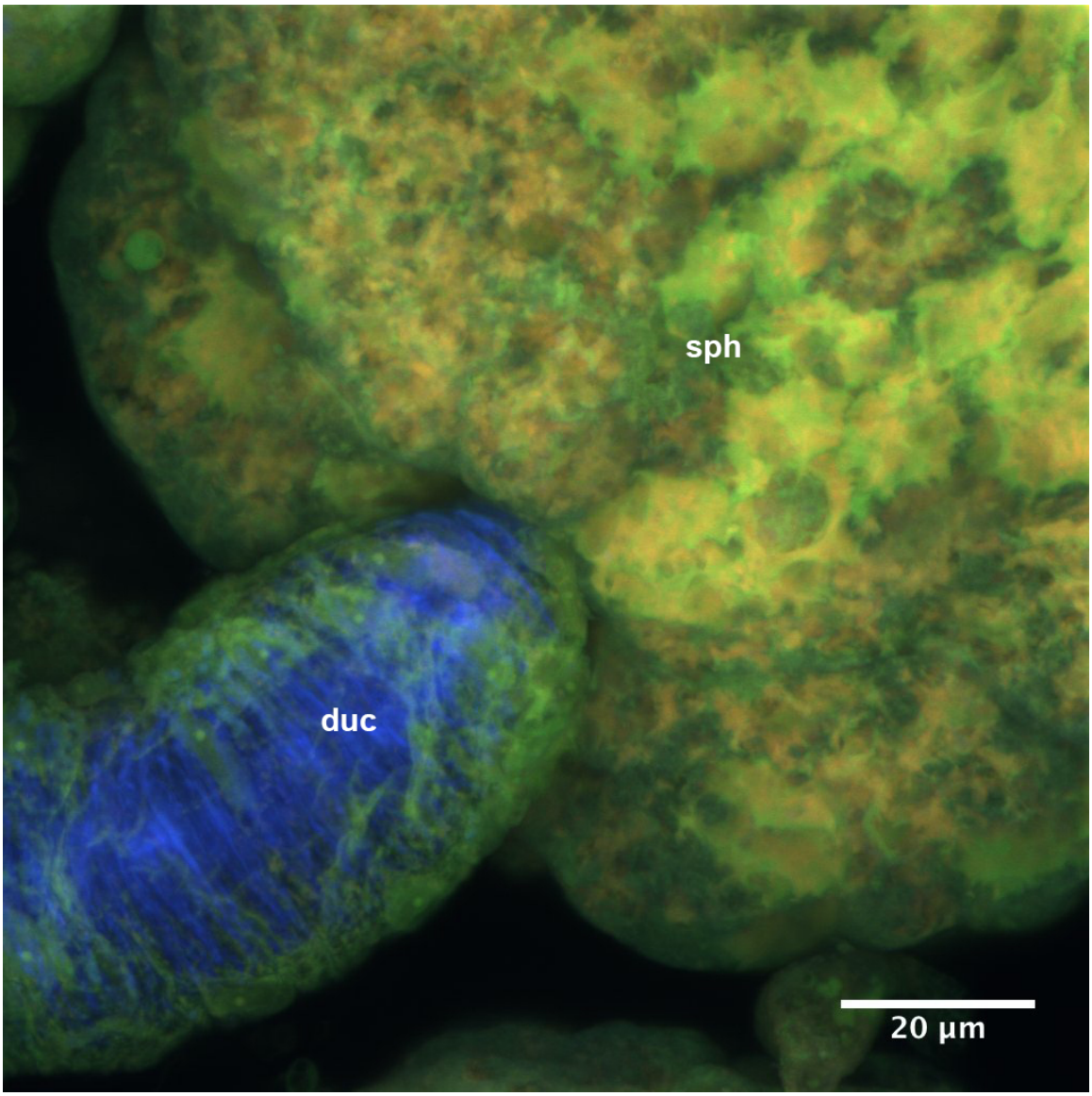


Figure 2. Spherical and collecting duct of secretory lobe. CLSM volume rendered micrograph showing insertion of a ductule (duc) into a secretory lobe spherical (sph) of *Harpalus pensylvanicus*. The ductule of the secretory lobe spherical is showing blue autofluorescence while the glandular sphere autofluoresces in green/orange. The ductule is covered by a few µm thick, green-autofluorescing epithelial cell layer.

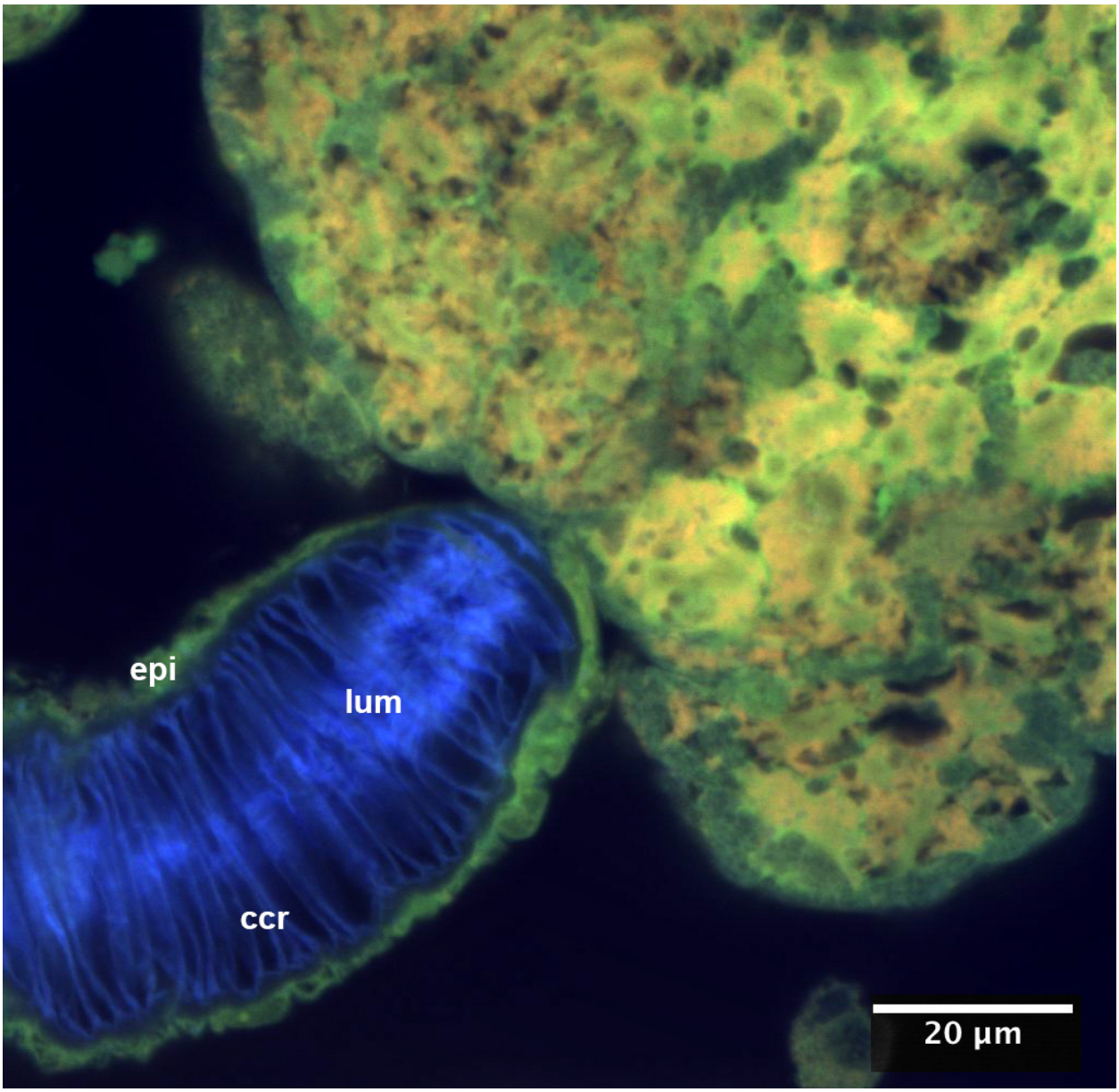


Figure 3. Sections of secretory lobe showing duct lumen, epithelial layer. CLSM optical section of a *Harpalus pensylvanicus* secretory lobe glandular sphere and ductule. The 3-6 µm thick lumen wall of the ductule (lum) is connected to the few µm thick, green autofluorescing ductile epithelium (epi) by the lamellar cuticular crista (ccr).

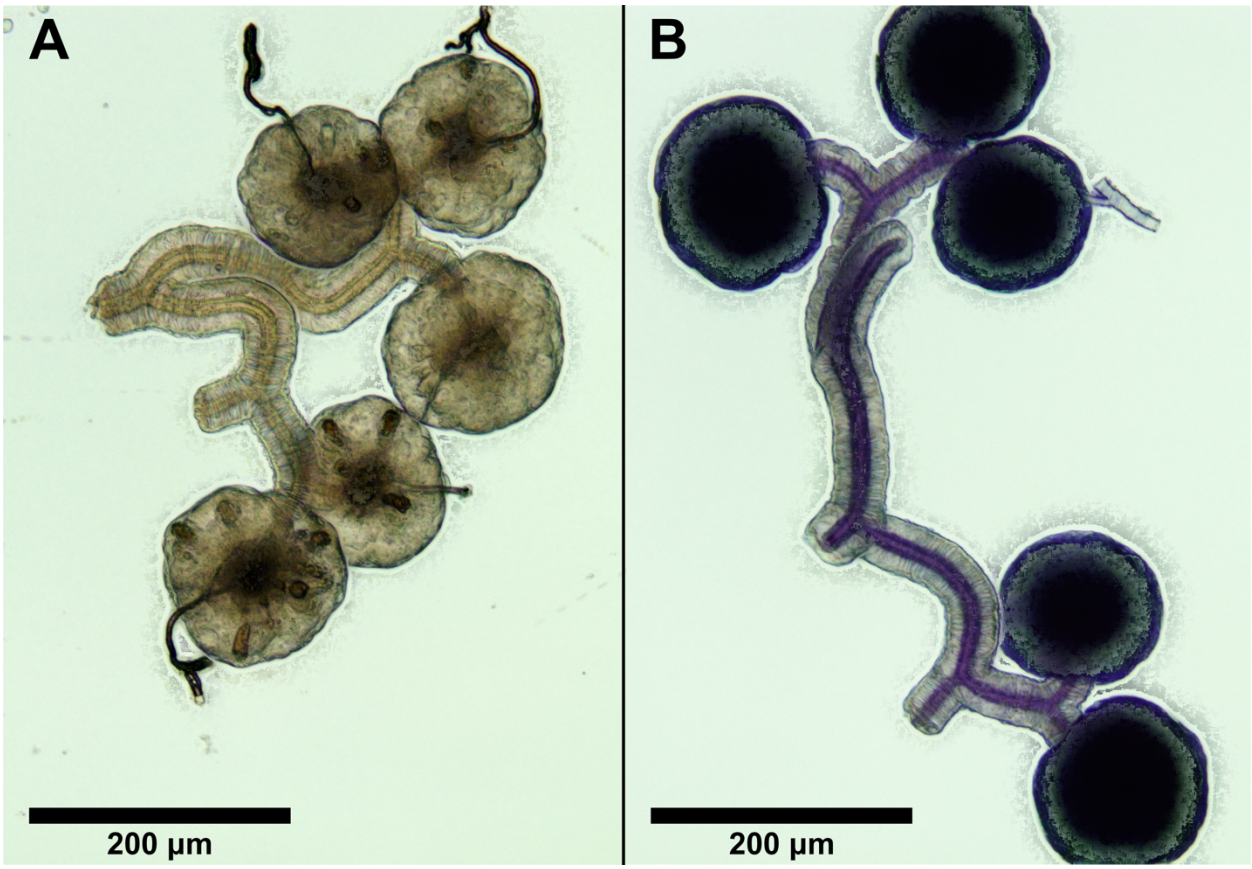


Figure 4a,b. Toluidine blue staining of a secretory lobe section. Brightfield image of a *Harpalus pensylvanicus* secretory lobe section before (left, A) and after staining for eighteen hours with 0.005% toluidine blue solution (right, B). After washing for five hours in 0.1M sodium phosphate buffer, the glandular spheres and the collecting duct lumen remained stained. The collecting duct lumen stained without showing metachromasia as has been described in other resilin-containing tissues.

CHAPTER 3

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

This work has been previously published and has been reproduced with permission from Springer Nature (https://doi.org/10.1007/s10886-021-01253-2).

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-deterring 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 (mTHF) 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.

INTRODUCTION

Arthropods are unique among animals in the diversity and novelty of chemical defensive strategies employed (Roth & 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 (Roth & Eisner 1962, Rork et al. 2019, Forsyth 1972, Will et al. 2000, Arndt et al. 2016 Attygalle et al. 2020, Eisner et al. 2006). 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 & Parker 2020).

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 (Rork & Rennner 2018, Lečić et al. 2014, Moore & Wallbank 1968, Attygalle et al. 2020). Even aquatic relatives of the Carabidae, the hydradephagans which include whirligig and predaceous diving beetles, produce various sesquiterpenes, norsesquiterpenes, aliphatic alcohols, and aldehydes (Ivarsson et al. 1996, Francke & Dettner 2005). 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 & Holoubek 1961).

One of the most common ground beetle defensive compounds is formic acid, a metabolite ubiquitous across life (Kanehisa & Murase 1977, Moore & 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) (Will et al. 2000, Hefetz & Blum 1978, Attygalle et al. 1993). 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) (Will et al. 2010, Wilson & Regnier 1971, Roth & Eisner 1962). 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 (Meiser et al. 2016, Brosnan & Brosnan 2016). 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 linages, 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 & 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 & 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 x 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 minutes and increased at 15 °C/min to a final temperature of 240 °C and held for 3 minutes. 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 (eventual NCBI SRA and TSA accessions will go here).

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 (150bp, paired-end) at a target depth aim of 50M 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 preprocessing 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, Hass et al. 2013). 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 including *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 (Hass 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 (Hass 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 & Eddy 2013). SignalP and TMHMM were used to predict potential coding sequences and transmembrane helices within amino acid sequences respectively (Petersen et al. 2011, Krogh et al. 2001). 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 (Kanehisa et al. 2012, Jensen 2008, 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 (Ritchie et al. 2015, Law et al. 2014). 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 & 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 & 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 & 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 & 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 1,000 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 Figure 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 & Kawazu 1982, Kanehisa & Kawazu 1985, McCullough 1966, Kanehisa & Murase 1977, Schildknecht et al. 1961c, Schildknecht 1964, Schildknecht et al. 1968a,c,d). This is also the eighth *Harpalus* species known to produce 2-tridecanone (Kanehisa & 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) (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 1,704 bp while the E90N50 was similar at 1,838 bp (Table 2). Querying the *H. pensylvanicus* transcripts against the endopterygota_odb9 database revealed a 94.1% complete transcriptome (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 (Table 3). 36.7% of transcripts received at least one BLASTX hit with an e-value < 1e-5 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 (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, 3a). 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. 3a, Supplementary Figure 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. 3a, Supplementary Figure 3).

Lipophilic Molecule Biosynthesis

Fatty-acyl CoA reductases 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, monocarboxylate transporters (MOTs), and proton-couples 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 (PATHs) which may serve as means by which the precursors of 5,10-methylenetetrahyrofolate, serine and glycine, 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 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 (Quistgaard et al. 2016, Locher 2009). 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 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 "methenyltetrahydrofolate 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 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 they are a tissue which utilizes the aforementioned pathway for formic acid biosynthesis. The abundance of enriched GO terms for organic acid and organic anion transport suggest that these compounds are not being used for endogenous metabolism so much as they are being secreted to the extracellular space (e.g. the collecting duct). Clearly 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 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 *FolD,* which contains both THF/DHG/CYC catalytic and NAD(P)$^+$-binding domains, duplicated in a common ancestor of Amorphea (Amoebozoa + Holozoa + Nucletmycea) (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) (Philippe et al. 2005, Küke et al. 2012).

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 nucletmycean 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 nucletmyceans, 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) (Philippe et al. 2005, Küke et al. 2012). 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 *MTHFD1*a-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 & Wallbank 1968, Rork et al. 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 & 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 Psydrinae) (Kanehisa & Kawazu 1982, Kanehisa & Kawazu 1985, Kanehisa & Murase 1977, Moore & Wallbank 1968, Rossini et al. 1977, Attygalle & Will, unpublished data). *H. pensylvanicus* also produces 2-tridecanone, as well as undecane, an alkane found in the sprays of other formic acid-producing Carabidae (Kanehisa & 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 & 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 Psydrinae 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 & 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) (Hartman & Buchanan 1959, Carreras & Santi 1995, Fox & Stover 2008). 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 (Mehler & Knox 1950, Han et al. 2012). 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 (Jaffe & Chrin 1979, Froese et al. 2019, 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,b). 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 glandless 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 (Scott & Wen 2001, Coppin et al. 2012, Jackson et al. 2013, Schuler & Berenbaum 2013). 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 & 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 (Li & White 2003, Wray 2007, Smith et al. 2013, Arsovski et al. 2015, Long et al. 2016, Sonawane et al. 2017, Steuernagel et al. 2019). 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 (Carroll et al. 2009, Ben-Sahra et al. 2016). 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 & Blum 1978a,b). 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.

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]. The findings and conclusions of this work do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. Journal of Molecular Biology. 215: 403-410.
2. Arndt E, Beutel RG, and 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). (pp. 119–144). Walter de Gruyter, Berlin
3. 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 Physiology. 169(4): 2982-2091.
4. Attygalle AB, Smedley SR, Meinwald J, Eisner T. (1993). Defensive secretion of two notodontid caterpillars (Schizura unicornis, S. badia). Journal of Chemical Ecology. 19(10): 2089-2104
5. Attygalle AB, Wu X, and Will KW. (2006). Biosynthesis of Tiglic, Ethacrylic, and 2-Methylbutyric Acids in a Carabid Beetle, *Pterostichus (Hypherpes) californicus*. Journal of Chemical Ecology. 33(5): 963-970.
6. 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*. The Science of Nature 107: 26.
7. Badawy AA. (2017). Kynurenine Pathway of Tryptophan Metabolism: Regulatory and Functional Aspects. International Journal of Tryptophan Research. 10: 1-20.

8. 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
9. Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. 57(1): 289–300
10. Bray NL, Pimentel H, Melsted P, Pachter L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology. 34: 525-527.
11. Brosnan ME & Brosnan JT. (2016). Formate: The neglected member of one-carbon metabolism. Annual Review of Nutrition. 36: 369-388.
12. Brückner A, Parker J. (2020). Molecular evolution of gland cell types and chemical interactions in animals. Journal of Experimental Biology. 223: jeb211938.
13. Carreras CW, Santi DV. (1995). The catalytic mechanism and structure of thymidylate synthase. Annual Review of Biochemistry. 64: 721-762.
14. 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. Human Genetics. 125(3):247-56.
15. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17: 540-552.
16. Chen S, Zhou Y, Taru C, Gu J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34: i884-i890

17. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. (2016). A survey of best practices for RNA-seq data analysis. Genome Biology. 17: 13.
18. 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 Biochemistry and Molecular Biology. 42(5): 343-352.
19. Deuve T. (1993). L'abdomen et les genitalia des femelles de Coleopteres Adephaga. Memoires du Museum National D'Histoire Naturelle. 155: 1-184.
20. Di Pietro E, Sirois J, Tremblay ML, MacKenzie RE. (2002). Mitochondrial NAD-Dependent Methylenetetrahydrofolate Dehydrogenase-Methenyltetrahydrofolate Cyclohydrolase Is Essential for Embryonic Development. Molecular and Cell Biology. 22(12): 4158-4166.
21. Eddy SR. (1998). Profile hidden Markov models. Bioinformatics. 14(9): 755-763.
22. 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.
23. 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). Journal of Insect Physiology. 6(4): 272-292.

24. Feyereisen R. (2020). Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Molecular Phylogenetics and Evolution. 143: 106695.
25. Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Research. 39(Web Server issue): W29-W37.
26. Forsyth DJ. (1972). The structure of the pygidial defence glands of Carabidae (Coleoptera). Transactions of the Zoological Society of London. 32: 249-309.
27. Forsyth, DJ. (1968). The structure of the defence glands in the Dytiscidae, Noteridae, Haliplidae and Gyrinidae (Coleoptera). Transactions of the Royal Entomological Society of London. 120: 158-181.
28. Forsyth, DJ. (1970). The structure of the defence glands of the Cicindelidae, Amphizoidae, and Hygrobiidae (Insecta: Coleoptera). Journal of Zoology, London. 160: 51-69.
29. Fox JT, Stover PJ. (2008). Folate-mediated one-carbon metabolism. Vitamins and Hormones. 79: 1-44.
30. Froese DS, Fowler B, Baumgartner MR. (2019). Vitamin B12, folate, and the methionine remethylation cycle- biochemistry, pathways, and regulation. Journal of Inherited Metabolic Diseases. 42(4): 673-685.
31. 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. Nature Biotechnology. 29(7): 644-652.

32. 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 Research. 77(4): 937-948.

33. 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. Nature Protocols. 8(8): 1494-1512.

34. Han Q, Robinson H, Li J. (2012). Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. Biochemical Journal. 446(2): 253-260.

35. 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.

36. Hartman SC, Buchanan JM (1959). Biosynthesis of the purines. XXVI. The identification of the formyl donors of the transformylation reactions. The Journal of Biological Chemistry. 234(7): 1812–1816

37. Hefetz A, Blum MS. (1978). Biosynthesis and accumulation of formic acid in the poison gland of the carpenter ant *Camponotus pensylvanicus*. Science. 201(4354): 454-455.

38. Hefetz A, Blum MS. (1978). Biosynthesis of formic acid by the poison glands of formicine ants. Biochimica et Biophysica Acta. 543(4): 484-496

39. Ivarsson P, Henrikson BI, Stenson JAE. (1996). Volatile substances in the pygidial secretion of gyrinid beetles (Coleoptera: Gyrinidae). Chemoecology. 7(4): 191-193.

40. 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 (αEsterase7) associated with insecticide resistance. Proceedings of the National Academy of Sciences of the United States of America. 110(25): 10177-10182.

41. Jaffe JJ, Chrin LR. (1979). De novo syntesis of methionine in normal and brugia-infected Aedes aegypti. The Journal of Parasitology. 65(4): 550-554.

42. 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 Research. 36(Database issue): D250-254.

43. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods.14(6): 587-589.

44. 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. Molecular Brains. 2: 8.

45. Kanehisa K, Kawazu K. (1982). Fatty Acid Components of the defensive substances in acid-secreting carabid beetles. Applied Entomology and Zoology. 17(4): 460-466.

46. Kanehisa K, Kawazu K. (1985). Differences in neutral components of the defensive secretion in formic acid-secreting carabid beetles. Applied Entomology and Zoology. 20(3): 299-304.

47. Kanehisa K, Murase M. (1977). Comparative study of the pygidial defensive systems of carabid beetles. Applied Entomology and Zoology 12: 225-235.

48. Kanehisa M, Goto S, Sato Y, Furumichi M, and Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular datasets. Nucleic Acids Research. 40: D109-D114.

49. Katoh K, Standley DM. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 30(4): 772-780.

50. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology. 305(3): 567-580

51. 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.

52. Law CW, Chen Y, Shi W, Smyth GK. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 15: R29.

53. Lečić S, Ćurčić S, Vujisić L, Ćurčić B, Ćurčić N, Nikolić Z, Anđelković B, Milosavljević S, Tešević V, Makarov S. (2014). Defensive secretions in three ground-beetle species (Insecta: Coleoptera: Carabidae). Annales Zoologici Fennici. 51: 285-300.

54. Li T, White KP. (2003). Tissue-specific gene expression and ecdysone-regulated genomic networks in *Drosophila*. Developmental Cell. 5(1): 59-72.

55. Locher KP. (2009). Structure and mechanism of ATP-binding cassette transporters. Philosophical Transactions of the Royal Society B: Biological Sciences. 2364(1514): 239-245.

56. Long HK, Prescott SL, Wysocka J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell. 167(5): 1170-1187.

57. MacFarlane AJ, Perry CA, Girnary 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. Journal of Biological Chemistry. 284(3): 1533-1539.

58. McCullough T. (1966). Compounds Found in the Defensive Scent Glands of *Harpalus caliginosus* (Coleoptera: Carabidae). Annals of the Entomological Society of America. 59(5): 1020-1021.

59. Mehler AH, Knox WE. (1950). The conversion of tryptophan to kynurenine in liver. II. The enzymatic hydrolysis of formylkynurenine. Journal of Biological Chemistry. 187(1): 431-438.

60. Meiser J, Schuster A, Pietzke M, Vande Voorde J, Athineos D, Oizel K, Burgos-Barragan G, Wit N, Dhayade S, Morton JP, Dornier E, Sumpton D, Mackay GM, Blyth K, Patel KJ, Niclou SP, Vazquez A. (2018). Increased formate overflow is a hallmark of oxidative cancer. Nature Communications. 9(1): 1368.

61. Meiser J, Tumanov S, Maddocks O, Labuschagne CF, Athineos D, Van Den Broek N, Mackay GM, Gottlieb E, Blyth K, Vousden K, Kamphorst JJ, Vazquez A. (2016).

Serine one-carbon catabolism with formate overflow. Scientific Advances. 2(10): e1601273.

62. Moore BP, Wallbank BE. (1968). Chemical composition of the defensive secretion in carabid beetles and its importance as a taxonomic character. Proceedings of the Royal Entomological Society of London. 37(5-6): 62-72.
63. Moore BP. (1979). Chemical Defense in Carabids and Its Bearing on Phylogeny. Carabid Beetles, 193–203.
64. Mulder GJ. (1992). Glucuronidation and its Role in Regulation of Biological Activity of Drugs. Annual Review of Pharmacology and Toxicology. 32: 25-49
65. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 32(1): 268-274.
66. 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.
67. Peschke K Eisner T. (1987). Defensive secretion of the tenebrionid beetle *Blaps mucronata* physical and chemical determinants of effectiveness. Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology. 161(3): 377–388.
68. Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods. 8: 785-786.
69. Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. (2005). Heterotachy and long-branch attraction in phylogenetics. BMC Evolutionary Biology. 5: 50.

70. Punta M, Coggill 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 Research. 40: D290-301.

71. 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. Proceedings of the National Academy of Sciences of the United States of America. 109(37): 14858-1463.

72. Quistgaard EM, Löw C, Guettou F, Nordlund P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nature Reviews Molecular Cell Biology. 17: 123-132.

73. 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.

74. 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 Research. 43(7): e47.

75. Rork AM, Mikó I, Renner T. (2019). Pygidial glands of *Harpalus pensylvanicus* (Coleoptera: Carabidae) contain resilin-rich structures. Arthropod Structure & Development. 44(12): 1069-1083.

76. Rork AM, Renner T. (2018). Carabidae Semiochemistry: Current and Future Directions. Journal of Chemical Ecology. 49: 19-25.

77. 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*). Proceedings of the National Academy of Sciences of the United States of America. 94: 6792-6797.

78. Roth LM, Eisner T. (1962). Chemical Defenses of Arthropods. Annual Review of Entomology. 7: 107-136.

79. 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. Biochemical Pharmacology. 34(3): 361-367.

80. Schildknecht H, Holoubek K. (1961). Die Bombardierkäfer und ihre Explosionschemie V. Mitteilung über Insekten-Abwehrstoffe. Angewandte Chemie. 73(1): 1-7.

81. Schildknecht H, Maschwitz U, Winkler H. (1968). Zur Evolution der Carabiden-Wehrdrüsensekrete. Über Arthropoden-Abwehrstoffe XXXII. Naturwissenschaften. 55: 112–117.

82. Schildknecht H, Weis KH. (1961). Die chemische natur des wehrsekretes von *Pseudophonus* pubescens und *Ps. griseus*. VIII. Mitteilung über insektenabwehrstoffe. Zeitschrift für Naturforschung B. 16: 361-363.

83. Schildknecht H, Winkler H, Maschwitz U. (1968). Vergleichend chemische Untersuchungen der Inhaltsstoffe der Pygidialwehrblasen von Carabiden. Über Arthropoden-Abwehrstoffe XXXI. Zeitschrift für Naturforschung B. 23: 637.

84. Schildknecht H. (1964). Defensive substances of the Arthropods, their isolation and identification. Angewandte Chemie International Edition. 3(2): 73-82

85. Schildknecht H. (1970). The Defensive Chemistry of Land and Water Beetles. Angewandte Chemie International Edition. 9(1): 1-9.

86. Schuler MA, Berenbaum MR. (2013). Structure and function of cytochrome P450s in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. Journal of Chemical Ecology. 39(9): 1232-1245.

87. 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.

88. Scott JG, Wen Z. (2001). Cytochromes P450 of insects: the tip of the iceberg. Pest Management Science. 57(10): 958-967.

89. Sekowska A, Ashida H, Danchin A. (2019). Revisiting the methionine salvage pathway and its paralogues. Microbial Biotechnology. 12(1): 77-97.

90. 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.

91. 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. Molecular Biology and Evolution. 30(11): 2509–2518.

92. 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. Molecular Biology Evolution. 32(5): 1342-1353.

93. 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 Reports. 21(4): 1077-1088.
94. 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.
95. Storey JD. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B. 64: 479–498.
96. Talavera G, Castresana J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Systematic Biology 56: 564-577.
97. Tautz D, Domazet-Lošo T. (2011). The evolutionary origin of orphan genes. Nature Reviews Genetics. 12(10): 692-702.
98. The Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics. 25: 25-29.
99. 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.
100. West MG, Barlowe CK, Appling DR. (1993). Cloning and characterization of the Saccharomyces cerevisiae gene encoding NAD-dependent 5,10-

methylenetetrahydrofolate dehydrogenase. Journal of Biological Chemistry. 268(1): 153-160.

101. Wheeler TJ, Eddy SR. (2013). nhmmer: DNA homology search with profile HMMs. Bioinformatics. 29(19): 2487-2489

102. Will KW, Attygalle AB, Herath K. (2000). New defensive chemical data for ground beetles (Coleoptera: Carabidae): interpretations in a phylogenetic framework. Biological Journal of the Linnean Society. 71: 459-481.

103. 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. Journal of Insect Science. 10: 1-17.

104. Wilson EO, Regnier FE. (1971). The evolution of the alarm-defense system in the formicine ants. The American Naturalist. 105(943): 279-289.

105. Wray GA. (2007). The evolutionary significance of cis-regulatory mutations. Nature Reviews Genetics. 8: 206-216.

106. Young MD, Wakefield MJ, Smyth GK, Oshlack A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 11: R14.

FIGURES AND TABLES

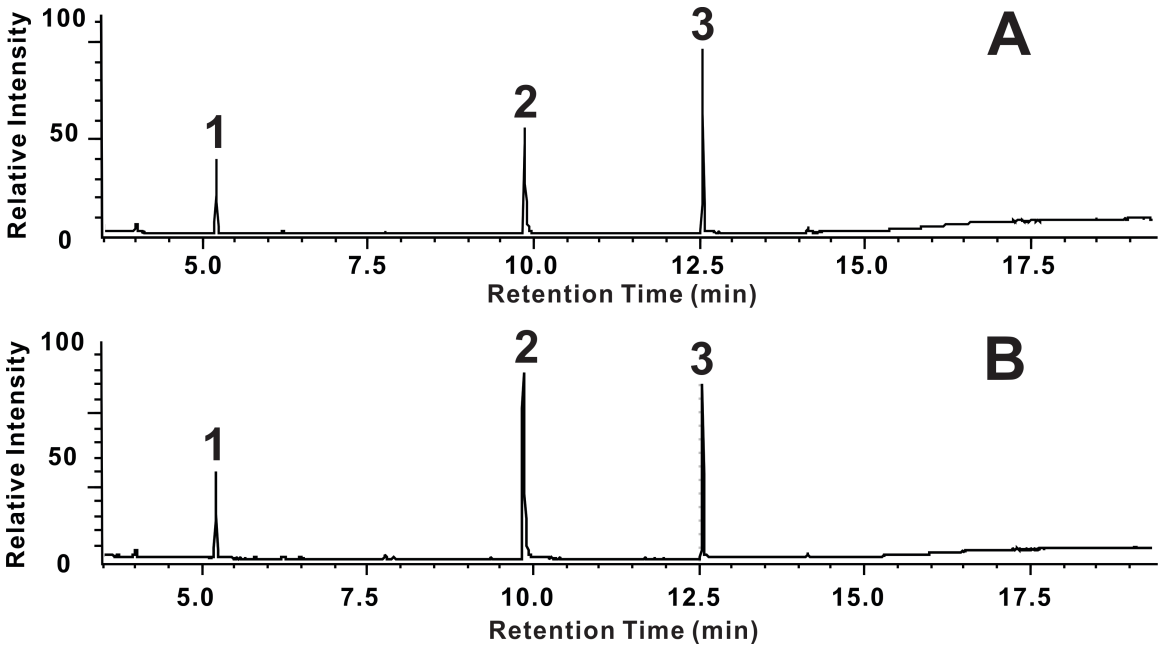


Figure 1a,b. Gas chromatograms generated via GC-MS analysis of the pygidial gland sprays of a male (A) and female (B) *Harpalus pensylvanicus* specimens into methylene chloride. Peak 1: Undecane; Peak 2: Formic acid; Peak 3: 2-Tridecanone. All samples run on a ZB-Wax column.

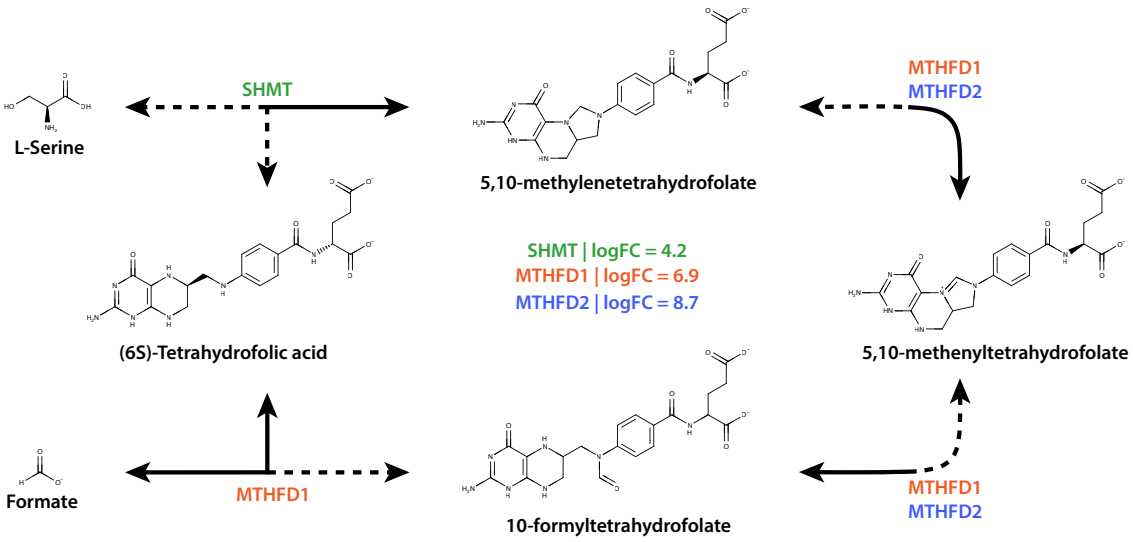


Figure 2. Simplified representation of a portion of the folate cycle of one-carbon metabolism in insects. Solid arrows represent an oxidative direction of pathway flow where the primary output is formate and tetrahydrofolate. Dashed arrows represent a reductive direction of pathway flow where L-serine 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 ,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.

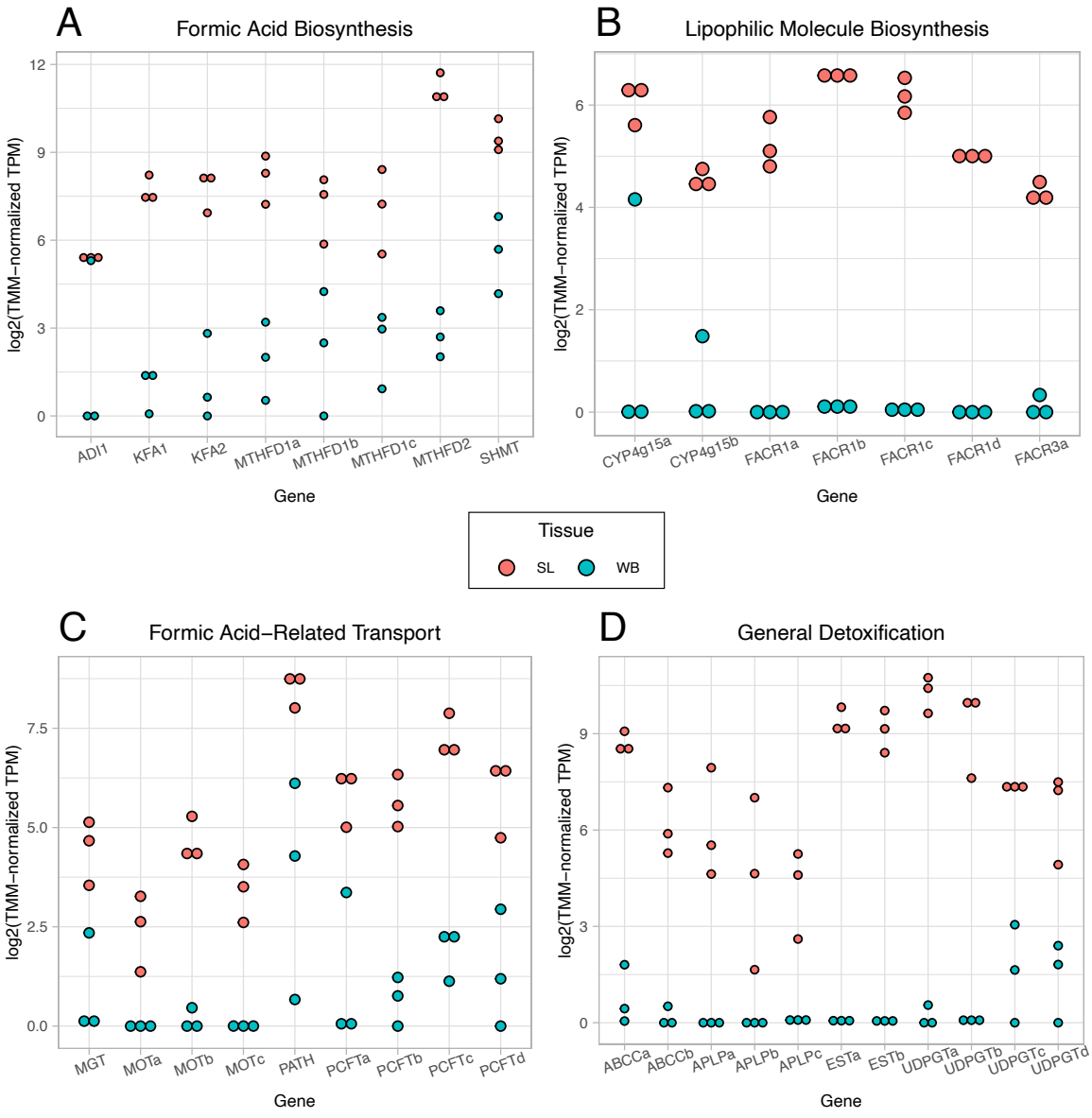


Figure 3a-d. Dot plots demonstrating the expression levels of genes discussed as part of this study in TMM-normalized TPM. Gene names follow the naming convention used throughout the manuscript with multiple copies/gene family members being denoted by

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.

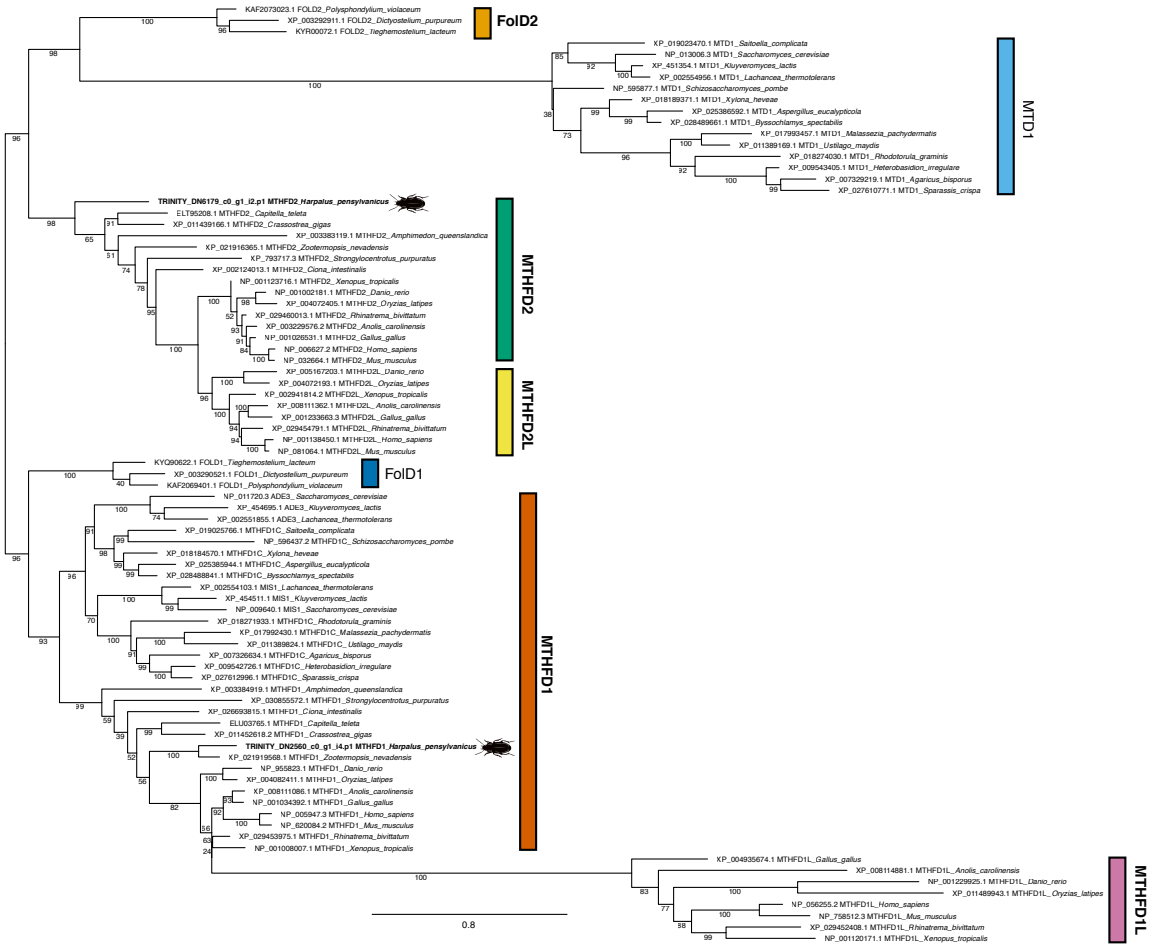


Figure 4. Maximum-likelihood phylogeny of the mTHF gene family with branch support values shown. Phylogeny was generated from a protein alignment of the THF/DHG/CYC catalytic and NAD(P)+-binding domains of mTHF gene family members across Amorphea using IQTREE v1.6.12. Branch support values were generated via ultrafast bootstrap approximation (1,000 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 Nucletmycea-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 clade 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.

Overview of the MTHFD Gene Family in Amorpheans

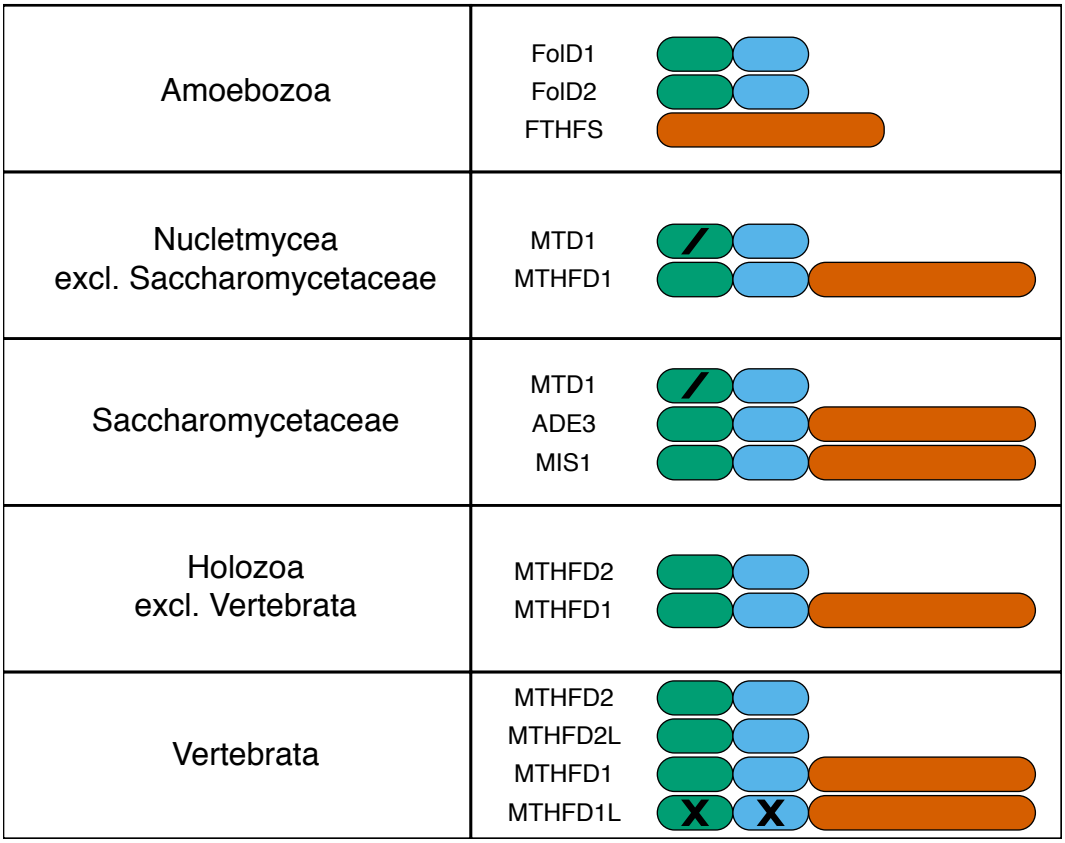


Figure 5. Overview of basic mTHF gene family architecture in amorphean taxa. The THF/DHG/CYC catalytic and NAD(P)$^{+}$-binding domains are shown in green and blue respectively. The formyltetrahydrofolate synthetase domain is depicted in orange. A

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 mTHF gene family evolution.

Sequencing & Quality Control Statistics	
Category	**Value**
Raw reads	300,049,838
Raw bases	45,307,525,538
Reads post-trimming	277,523,094
Bases post-trimming	41,101,307,864
Reads too short	14,978,184
Reads low quality	6,851,582
Reads low complexity	692,208

Table 1. Statistics on raw sequencing data output, data retained post-trimming, and data discarded via trimming. Length cutoff was set at 100bp, quality cutoff at Q20, and complexity cutoff at 30% complexity.

Transcriptome Assembly Statistics	
Category	**Value**
Total No. Genes	94,288
Total No. Transcripts	160,301
Total Bases Assembled	165,136,649
Mean Transcript Length (bp)	1,030
Transcript N50 (bp)	1,704
Transcript E90N50 (bp)	1,838
Complete BUSCOs	2103 (94.1%)
Fragmented BUSCOs	68 (2.8%)
Missing BUSCOs	76 (3.1%)

Table 2. Transcriptome assembly and quality control statistics (reported as BUSCO v3 scores using the endopterygota database).

Transcriptome Annotation Statistics	
Category	**Value**
Total Transcripts	160,301
Predicted Peptide Sequences	70,782 (44.2%)
Transcripts with 1+ BLASTX hit	58,875 (36.7%)
Peptides with 1+ BLASTP hit	50,049 (70.7%)
Peptides with 1+ HMMScan hit	47,156 (66.6%)
Peptides with GO Annotations	57,777 (81.6%)

Table 3. Transcriptome annotation statistics. BLASTX and BLASTP hits were against the UniProtKB database. HMMScan hits were against the Pfam-A database.

CHAPTER 4

Comparative Transcriptomic Analyses Unveil the Molecular Mechanisms of Ground Beetle Chemical Defense

ABSTRACT

From the kairomones that parasitoids use to eavesdrop on hosts, to the sex pheromones used by moths for mate attraction, insect semiochemicals have been of great interest to the chemical ecology community for decades. Indeed, their utility in honing our understanding of insect ecology and evolution cannot be understated. One such group that has been the topic of substantial focus over the past several decades are the Carabidae, namely due to the diversity of defensive compounds they synthesize and secrete. Over 500 species' defensive chemicals have been characterized to date, culminating in a library of hundreds of unique compounds from caustic acids to pungent phenolics to distasteful sulfides, and in some cases even citrusy, piney terpenes. However, despite our breadth of understanding of Carabidae allomone chemistry, our knowledge of semiochemical evolution within the family and how they are biosynthesized is limited. To date, only four studies have demonstrated the likely biosynthetic precursors for a small number of compounds and only one has demonstrated the genes that may be involved in the biosynthesis of formic acid. Here, we aim to expand upon previous work by reporting on the defensive gland transcriptomes of six species spanning three subfamilies of Carabidae and three chemical classes. We include the Harpalinae formic acid producers *Harpalus pensylvanicus* and *Platynus angustatus*, as well as the methacrylic acid-producing harpaline *Pterostichus moestus*. Our

benzoquinone producers from the Brachininae are *Brachinus elongatulus* and *Pheropsophus janus*, whereas the benzoquinone producer from the Paussinae is *Goniotropis kuntzeni*. We demonstrate the likely role that the folate cycle of one-carbon metabolism plays in the biosynthesis of formic acid from L-serine, and the role that valine catabolism likely plays in converting L-valine to methacrylic acid. We also demonstrate the presence of peroxidases in the reaction chambers of our quinone producers, which likely play a role in the oxidation of hydroquinones to benzoquinones. It remains uncertain how hydroquinones themselves may be biosynthesized from simpler phenolic precursors, although phenoloxidases and reductases likely play a substantial role. This work provides additional evidence that Carabidae have repeatedly co-opted primary metabolic pathways for defensive chemical biosynthesis in their defensive glands and also provides evidence that benzoquinones may be biosynthesized through a yet undescribed mechanism.

INTRODUCTION

Arthropods are well-known for not only their taxonomic diversity, but their biochemical diversity (Roth & Eisner 1962). For many species, biochemicals are the sole or primary means of communicating ecologically relevant information with conspecific and heterospecifics alike. Eusocial insects such as ants emit trail pheromones to inform nestmates about the location of resources (Attygalle & Morgan 1985). Male Lepidoptera emit sex pheromones from coremata to advertise their location to potential mates as well as repel competitors (Birch et al. 1990). Bark beetles of the subfamily Scolytinae are well known for emitting aggregation pheromones that attract swarms of conspecifics to trees for oviposition, resulting in vast quantities of larvae feeding upon the tree and overwhelming induced defenses (Symonds & Elgar 2003). Millipedes secrete toxic quinones, cyanide, and benzaldehyde in response to perturbation by organisms ignoring their aposematism (Rodriguez et al. 2018). These semiochemicals have been of interest to the scientific community for decades given their potential to explain patterns of insect evolution, behavior, and even have practical purposes as pest management tools. Amongst the most well studies are insect defensive compounds, which include small organic molecules, venoms, and complex polymers alike (Attygalle et al. 1993, Eisner et al. 1961, Roth & Eisner 1962, Laxme et al. 2019). The former comprise the largest area of inquiry and include a wide variety of metabolites such as carboxylic acids, quinones, phenolics, terpenes, sulfides, nitro-containing compounds, etc. (Roth & Einser 1962, Moore & Wallbank 1968). Indeed, hundreds if not thousands of small organic defensive compounds have been described across Arthropoda, mostly from exocrine secretions. Despite our depth of knowledge regarding the diversity of insect defensive compounds,

our knowledge of their biosynthesis is wanting. Great strides have been made in recent years throughout Arthropoda aimed at filling this gap in knowledge, including in one of the phylum's most diverse lineages, the Adephaga.

The Adephaga are a hyperdiverse suborder of beetles comprised of 45,000 species spanning 11 families (Arndt et al. 2016, Vasilikopoulos et al 2021). Traditionally Adephaga has be further subdivided into two groups based on life history: the 'Hydradephaga' and the Geadephaga (Crowson 1960, Gustafson et al. 2019, Vasilikopoulos et al 2021). The 'Hydradephaga' represent 8 families within Adephaga, but only a small share of the total species diversity (5,000 spp.). These include charismatic groups such as the predaceous diving beetles (Dytiscidae) and whirligig beetles (Gyrinidae), all of which are aquatic. However, Hydradephaga have been shown in recent years to be paraphyletic without the inclusion of Geadephaga (Gustafson et al. 2019, McKenna et al. 2019, Vasilikopoulos et al 2021). The Geadephaga, while representing only 3 of the 11 families, comprise most of the species (~40,000 spp.). Of these, most geadephagan species can be found in a single family, the Carabidae, followed by the tiger beetles (Cicindelidae) and the false ground beetles (Trachypachidae). Nearly half of all adephagans belong to a single subfamily of Carabidae, the Harpalinae (Bousquet 2012).

All Adephaga are united by, amongst other features, a unique pair of defensive glands situated in the abdomen aptly known as pygidial glands (Forsyth 1968, 1970, 1972, Deuve 1993). These glands are often comprised of four units, although some taxa have evolved additional gland features. These are the secretory lobes, the reservoirs, the collecting ducts which connect the two, and an efferent duct. Defensive chemicals are

thought to be biosynthesized in the secretory lobes, which are then transported through the resilin-rich collecting ducts to the reservoir for storage (Forsyth 1972, Rork et al. 2019, Muzzi et al. 2019). When a beetle is perturbed, the muscles of the reservoir contract, propelling defensive compounds out of the efferent duct which opens to the tip of the abdomen. In certain taxa, additional accessory glands have been found articulated to the efferent duct (Forsyth 1972). In others, the reservoir has been observed as divided into two lobes rather than the one typical of most taxa (Will et al. 2000). In two lineages, the bombardier beetles, the efferent duct has been modified into a sclerotized reaction chamber (Di Giulio et al. 2015, Muzzi et al. 2019). This reaction chamber secretes peroxidases which break down hydrogen peroxide introduced from the reservoir, generating heat, water, and molecular oxygen in the process (Schildknecht 1964, Aneshansley et al. 1969).

The defensive chemistry of adephagans is exceptionally diverse with well over two hundred chemicals having been described to date and likely several more left undiscovered (Rork & Renner 2018). These compounds can generally be broken down into larger chemical classes, including carboxylic acids, quinones, terpenes, sulfides, hydrogen cyanide, aromatic acids and aldehydes, hydrocarbons and their derivatives, etc. Of these, the most ubiquitous appear to be the carboxylic acids, namely methacrylic acid and formic acid (Schildknecht 1970, Kanehisa & Murase 1977, Kanehisa & Kawazu 1982, Kanehisa & Kawazu 1985, Will et al. 2000, Rork et al. 2019, Rork et al. 2021). Methacrylic acid is present in many major subfamilies of Adephaga, including the Carabinae, the Scaratinae, the Trechinae, and the Harpalinae (Moore & Wallbank 1968, Kanehisa & Murase 1977, Lečić et al. 2014). Formic acid has only been detected in the

Trechinae and Harpalinae but is especially widespread in the latter (Moore & Wallbank 1968, Kanehisa & Murase 1977, Moore 1979, Rossini et al. 1997, Rork et al. 2021). Also widespread are quinones, which while not being as abundant as formic or methacrylic acid, can be found in at least three subfamilies: the Paussinae, Brachininae, and Harpalinae (Moore & Wallbank 1968, Schildknecht et al. 1968, Kanehisa & Murase 1977). The latter appear to biosynthesize benzoquinones directly in the secretory lobes whereas the former two, the bombardier beetles, oxidize hydroquinones to benzoquinones in their reaction chambers. While often not discussed as primary defensive chemicals, aliphatic hydrocarbons, ketones, aldehydes, esters, and alcohols can be found in the pygidial gland secretions of most subfamilies (Moore 1979, Rossini et al 1997, Will et al. 2000).

Given the most current phylogenetic hypotheses for the family Carabidae, this pattern of defensive chemical presence across the family suggest that many compounds have evolved not once but several times independently (Vasilikopoulos et al 2021). Indeed, quinones are likely to have evolved at least three times, formic acid at least twice, terpenes twice, etc. (Moore & Brown 1979, Attygalle et al. 2009). Even the highly specialized bombardier beetle phenotype, which entails the biosynthesis of concentrated (>20%) hydrogen peroxide, has likely evolved twice in the family (Di Giulio et al. 2015, Muzzi et al. 2019, Vasilikopoulos et al 2021). This raises an interesting question about the nature of defensive chemical evolution in the Carabidae: in cases where independent lineages have evolved to secrete identical defensive chemicals, do they biosynthesize said compounds using identical or distinct biochemical pathways? For example, we have previously provided transcriptomic evidence that the formic acid producer *Harpalus*

pensylvanicus may be biosynthesizing its primary defensive compound via the folate cycle of C1 metabolism (Rork et al. 2021). However, there are several other mechanisms by which formic acid could be biosynthesized in Carabidae, including the kynurenine pathway, the methionine salvage cycle, alpha-oxidation of fatty acids, etc. (Meiser et al. 2016, Brosnan & Brosnan 2016, Badawy 2017). Perhaps this is the case not only for *H. pensylvanicus*, but for all formic acid-producing harpalines and even formic acid-producing trechinines. Alternatively, perhaps this is only the case in *Harpalus pensylvanicus* and other lineages use one of the other pathways mentioned or biosynthesize formic acid *de novo*. Unfortunately, aside from *H. pensylvanicus* and a few other taxa, we know very little about how carabid defensive chemicals are biosynthesized (Attygalle et al. 1991, Adachi et al. 1985, Attygalle et al. 2006, Attygalle et al 2020, Rork et al. 2021).

Here we aim to address this gap in knowledge by expanding upon previous work by providing transcriptomics-based data for six taxa spanning three defensive chemical phenotypes (or chemotypes) to identify candidate genes and pathways involved in formic acid (FA), methacrylic acid (MA), and benzoquinone (BQ) biosynthesis. These six taxa are: *Harpalus pensylvanicus*, *Platynus angustatus*, *Pterostichus moestus*, *Brachinus elongatulus*, *Pheropsophus jessoensis*, and *Goniotropis kuntzeni*. *H. pensylvanicus* and *P. angustatus* are formic acid producers of the subfamily Harpalinae. *P. moestus* is also of the subfamily Harpalinae but produces methacrylic acid. *B. elongatulus*, *P. jessoensis*, and *G. kuntzeni* are all benzoquinone-producing bombardier beetles, the former two being of the subfamily Brachininae, the latter from the subfamily Paussinae. We hypothesize that as was the case in *H. pensylvanicus*, *P. angustatus* secretory lobes will

also upregulate genes involved in the folate cycle of C1 metabolism, suggesting a role for it in the biosynthesis of defensive-grade formic acid. Previous work has shown that methacrylic acid is biosynthesized in the carabids *Carabus yaconinus* (Carabinae) and *Scarites subterraneus* (Scaratinae) from L-valine (Adachi et al. 1985, Attygalle et al. 1991). Although in different subfamilies, given this evidence from two independent taxa, we hypothesize that *P. moestus* is also upregulating genes in their secretory lobes involved in the larger branched-chain amino acid degradation pathway, as well as genes specific to the valine catabolic pathway. Isotope tracing data in *B. elongatulus* do not provide evidence that pathways commonly implicated in benzoquinone biosynthesis in other taxa are involved in the process here (Attygalle et al. 2020). Rather, it seems possible that the pathway has evolved *de novo*, perhaps through the co-option of existing enzymes. Nevertheless, we hypothesize that oxidoreductases or hydroxylases play an important role in producing 2-methyl-1,4-hydroquinone from its precursor, meta-cresol, and that peroxidases in the reaction chamber play a role in converting hydroquinones to their benzoquinone counterparts (Schildknecht 1964).

MATERIALS AND METHODS

Beetle collection

Harpalus pensylvanicus specimens were collected as described previously (Rork et al. 2021). Briefly, beetles collected from agricultural plots near Ramblewood, Pennsylvania between Sept. and Oct. 2017 were kept together in terrariums and were fed a diet of hulled millet with wet paper towels for water. *Platynus angustatus* and *Pterostichus moestus* specimens were collected at Stone Valley Recreation Area in Huntingdon

County, Pennsylvania throughout the summer and fall of 2019. Beetles were kept in individual plastic containers with coconut fiber substrate, were fed a diet of pecans and dog food (Castor & Pollux Organic Cookies, Chicken Recipe), and were provided with wet paper towels for water. *Brachinus elongatulus* were collected at Madera Creek in Madera Canyon, Arizona in April 2015. *Goniotropis kuntzeni* were collected from the Pinal Mountains in Pinal County, Arizona in September 2017. *Pheropsophus jessoensis* were collected in South Korea with the assistance of collaborators at Kyungpook National University, Sangju, South Korea.

GC-MS Analysis of Pygidial Gland Contents

Harpalus pensylvanicus gland exudates were analyzed as described previously (Rork et al. 2021). Hind legs of two male beetles were pinched, causing them to spray into 2-mL autosampler vials containing anhydrous dichloromethane. The same was done for two female beetles. Gland samples were analyzed on a Shimadzu GC-17A with a 30m x 0.25 mm x 0.25 μm ZB-WAX column coupled to a QP5050 mass spectrometer. Oven temperature was held at 40 °C for 3 minutes and increased at 15 °C/min. to a final temperature of 240 °C which was held for 3 minutes. Samples were introduced to the column by splitless injection, 1 μL/sample. *Platynus angustatus* and *Pterostichus moestus* were analyzed using the same instrumentation with the oven temperature held at 40 °C for 4 minutes and increased at 10 °C/min to a final temperature of 240 °C which was held for 5 minutes. Gland exudates were collected for both male and female *P. moestus*, although we were unable to determine the sex of the *P. angustatus* individuals sprayed. A typical external character used to determine sex is the size of the foretarsi, males often

having broader foretarsi than females. However, in *P. angustatus* this character was less apparent than in other taxa, making reliable identification difficult without a reference.

Brachinus elongatulus gland exudates were analyzed as described previously, by inducing beetles to spray onto small strips of filter paper by pinching hind legs with forceps, which were then placed into glass vials of hexane (Attygalle et al. 2020). Samples were analyzed on the same instrumentation, here instead with a 25 m x 0.20 mm x 1.12 μm DB-624 column. Oven temperature began at 100 °C and was increased by 10 °C/min to a final temperature of 250 °C. Samples were also analyzed via an Agilent HP 6890 GC with a 30m x 0.25 mm x 0.25 μm dimethylpolysiloxane column, coupled to an Agilent HP 5973 Mass Selective Detector (MSD). Oven temperature was held at 40 °C for two minutes and increased at 20 °C/min. to a final temperature of 280 °C, which was held for 7.5 minutes. In both cases, samples were introduced via splitless injection. *G. kuntzeni* was processed in a similar manner. The pygidial gland secretions of *P. jessoensis* have been characterized previously (Kanehisa & Murase 1977).

Pygidial Gland Dissections

Harpalus pensylvanicus specimens were dissected as described previously (Rork et al. 2021). Briefly, hind legs of male and female beetles were pinched to induce a defensive spray response. After waiting one hour, beetles were separated at the prothoracic-mesothoracic junction and were stored in RNAlater™ Stabilization Solution at -80 °C until dissection. Specimens were dissected under an Olympus SZX16 stereomiroscope using fine-tipped forceps in watch glasses filled with RNAlater. All instruments, surfaces, and tools were cleaned with ethanol and RNase Away Decontamination Reagent® prior to

dissection. Forceps were also flamed prior to RNase Away treatment. Beetles were dissected and pooled into two tissue types: secretory lobe tissue and whole-body tissue without secretory lobes. Wings and elytra were stored separately as vouchers. *Platynus angustatus* and *Pterostichus moestus* were processed similarly, although wait time between spraying and storing in RNAlater™ Stabilization Solution was reduced by half. Three biological replicates were collected for *H. pensylvanicus* and *P. angustatus* while two were collected for *P. moestus*.

Brachinus elongatulus were cut above the first abdominal segment and stored in RNAlater™ Stabilization Solution at 4 °C for no more than 48 hours prior to dissection. In addition to pooling secretory lobe tissue and whole-body tissue, the reactions chambers were also pooled, again in RNAlater™ Stabilization Solution. Three bioreplicates were generated for each tissue type. *P. jessoensis* and *G. kuntzeni* were processed similarly to *B. elongatulus*, although they were dissected the same day as being dispatched.

RNA Extractions, Library Prep, and Illumina Sequencing

Harpalus pensylvanicus samples were processed as described previously (Rork et al. 2021). Total RNA was extracted from secretory lobes using a modified TRIzol protocol and RNA isolated via a DirectZol RNA MicroPrep Kit. Total RNA from whole body tissue with secretory lobes removed was extracted and isolated via a TRIzol/chloroform/isopropanol protocol. *Platynus angustatus* and *Pterostichus moestus* total RNA was extracted from both secretory lobe samples and whole-body samples using a standard TRIzol protocol and isolated using a DirectZol RNA MiniPrep Kit.

Libraries were prepared for *H. pensylvanicus* using Illumina TruSeq and Nextera Library Preparation Kits while *P. angustatus* and *P. moestus* libraries were prepared using the latter. *H. pensylvanicus* libraries were sequenced on an Illumina NovaSeq 6000 (one lane, PE, 150 bp) at a target read depth of 50M/library. *P. angustatus* and *P. moestus* were also sequenced on an Illumina NovaSeq 6000 (two lanes, PE, 150 bp), but at target reads depths of 25-30M/library.

Brachinus elongatulus total RNA was extracted from each tissue type using an RNeasy Mini Kit and libraries prepared using a TruSeq RNA Sample Preparation Kit v2. *B. elongatulus* libraries were sequences on an Illumina HiSeq2000 (two lanes, PE, 100 bp). *Goniotropis kuntzeni* and *Pheropsophus jessoensis* samples were processed using a TRIzol/chloroform/isopropanol protocol combined with MicroBiospin-30 column (BioRad) nucleic acid purification. Libraries were prepared using a NEBNext Ultra II Directional RNA Library Prep Kit and were sequenced on an Illumina HiSeq4000 (one lane/species, PE, 150 bp).

Computing Resources and Bioinformatics Software

All bioinformatic analyses conducted as part of this study were carried out through the ACI-b (Advanced CyberInfrastructure - Batch) portion the Roar supercomputer (RHEL 7 OS) at The Pennsylvania State University unless otherwise noted. All bioinformatics software used was installed via miniconda3 unless otherwise noted.

Quality Assessment of Reads and Trimming

Prior to quality assessment, FASTQ files generated from libraries split across multiple lanes (i.e. *Brachinus elongatulus*, *Platynus angustatus*, and *Pterostichus moestus*) were concatenated sample-wise. For all six species, initial FASTQC (v0.11.9) reports were generated for each sample and were manually inspected to flag any major library quality issues. After initial quality assessment, adapter content for every library was examined using BBTools' (v38.90) BBDuk.sh script using the adapters.fa file as a reference database (Bushnell 2014). Predominant adapters were then trimmed using BBDuk.sh and the literal adapter strings as references. The 3' ends of reads were then trimmed to Q10 using the Phred algorithm. For species sequenced on two-color systems (*H. pensylvanicus, P. angustatus*, and *P. moestus*), poly-G tails were also trimmed. For all species, reads containing ambiguous bases (Ns) were discarded, as were those less than 75 bp in length post-trimming. FASTQC reports were subsequently generated for each trimmed library to ensure detectable adapters, N-containing reads, and short reads were adequately removed.

Transcriptome Assembly and Quality Assessment

Transcriptomes were assembled *de novo* using Trinity RNA-Seq (v2.12) (Grabherr et al. 2011, Haas et al. 2013). Default settings were used for each assembly with a few noted exceptions. Throughout this work, we do not consider thread usage or memory allocation when discussing default settings; in most cases, we adjust thread usage where applicable. Read orientation (--SS_lib_type) was specified as "RF" for each species except for *B. elongatulus* which was not sequenced using a strand-specific protocol. Minimum contig

length (--min_contig_length) was set to 300 bp for all species. SuperTranscripts were also generated for each species but are unused as of this present work (Davidson et al. 2017). Trimmomatic was left off given that reads had already been trimmed and *in silico* read normalization settings were left as default.

ExN50 values and transcript length statistics were generated using the "contig_ExN50_statistic.pl" and "TrinityStats.pl" scripts respectively. The "PtR" script was used to assess replicate-wise and sample-wise correlations within species. BUSCO was run on each transcriptome using default settings, the only exception being that analysis mode was set to "transcriptome" (Simão et al. 2015). The insect_odb9 database was used as the reference set of benchmark universal single-copy orthologs.

Read Quantification

Transcript quantification was carried out using the Kallisto (v0.46.2) pseudoalignment-based strategy via the "align_and_estimate_abundance.pl" script included as part of Trinity RNA-Seq's downstream analysis utilities (Bray et al. 2016). All settings were left as default. Transcripts and gene expression matrices were subsequently generated using the "abundance_estimates_to_matrix.pl" script, all settings again left as default.

Coding Sequence Prediction and Functional Annotation

Coding sequences were predicted *ab initio* from transcripts and were translated into their respective polypeptides via TransDecoder (v5.5.0) (Hass et al. 2013). To determine homology of assembled transcripts and their predicted proteins BLASTX and BLASTP (v2.11.0) searches were run against the UniProtKB database using the full set of

assembled transcripts and their predicted proteins respectively (Altschul et al. 1990). HMMScan (v3.3.2) searches were also run against the Pfam-A protein database using the full set of predicted proteins to assess protein domain content (Eddy 1998, Finn et al. 2011, Punta et al. 2012). Both the UniProtKB and Pfam databases were originally downloaded in March 2021 from the European Bioinformatics Institute's FTP site. At the time of analysis, Trinotate was unable to be utilized for loading functional annotation results into databases due to an incompatibility between links used to download annotations data (Hass et al. 2013). For this reason, a custom script was created to generate an analogous database structure, albeit without the typical eggNOG, KEGG, and GO information attached. A Trinotate database for each species was later generated after database synchronization problems were resolved (v3.2.2). For all BLAST and HMMScan searches, default search parameters were left as default except for the e-value threshold, which was set to 1e-10 (Altschul et al. 1990, Eddy 1998, Finn et al. 2011).

Differential Gene Expression Analyses and GO Enrichment

Kallisto quantification results were used as inputs to the "run_DE_analysis.pl" script, used here to conduct differential gene expression analyses between tissues of each species (Bray et al. 2016). For *Harpalus pensylvanicus, Platynus angustatus*, and *Pterostichus moestus*, contrasts were performed between secretory lobe (SL) samples and whole body (WB) samples in order to identify genes significantly upregulated in the secretory lobes relative to the whole bodies. For *Brachinus elongatulus*, *Goniotropis kuntzeni*, and *Pheropsophus jessoensis*, contrasts were also preformed between reaction chamber (RC) samples and whole body (WB) samples for the same reason. Voom was

used to log2-transform gene-wise count data to log-counts per million and limma (v3.42.0) was used to identify differentially expressed genes (Law et al. 2014, Ritchie et al. 2015). Genes were considered differentially expressed at a logFC of 4.0 or higher and a false discovery rate of 0.05 or below (Benjamini & Hochberg 1995, Storey 2002). Note that this logFC cutoff is rather high compared to the usual values of 1.5-2.0. This was done to only focus on the most upregulated genes in a tissue, those which are putatively most important for its function.

GO enrichment analyses were run using GOSeq (v1.38.0) via the run_goseq.pl script, which was modified such that all GO Terms would be printed rather than only those with an FDR less than 0.05 (Benjamini & Hochberg 1995, The Gene Ontology Consortium 2000, Young et al. 2010). Reports of enriched and depleted GO terms for each tissue type relative to the others was generated for each species. The go_basics.obo file needed to be downloaded manually, as the GO Consortium link changed between Trinity version releases (August 2021). Terms were classified as enriched at a false discovery rate of 0.05 or below. GO enrichment results were further summarized using the R package RRVGO (v1.4.4), which was not installed via miniconda3 (Sayols 2020).

Identification of Chemotype-specific Enriched GO Terms

To better assess which processes are important for chemical defense in a specific chemotype or lineage, we conducted multispecies combinatorial comparisons to identify which GO terms are commonly enriched in the secretory lobes of certain taxa. Specifically, we identified enriched GO terms exclusive to the secretory lobes of benzoquinone producers, exclusive to the secretory lobes of formic acid producers, and

exclusive to the secretory lobes of our methacrylic acid producer. This was done by piping the lists of enriched GO terms of each species into successive of grep commands, at each step filtering out terms found in non-target chemotypes, thus leaving only terms shared by members of a chemotype at the end of the filtering process. To generate our list of formic acid-specific GO terms, for example, we first file grepped the list of secretory lobe-enriched GO IDs of *Harpalus pensylvanicus* against the list of secretory lobe-enriched GO IDs of *Platynus angustatus*. This gave us a subset of GO Terms enriched in both species' secretory lobes. We then inverse grepped this list using the secretory lobe-enriched GO IDs of the other four species, effectively removing from the initial list any secretory lobe-enriched GO IDs shared between *H. pensylvanicus*, *P. angustatus*, and one or more of the other four taxa. This final list of GO IDs thus represented formic acid producer-specific secretory lobe-enriched GO IDs, which were then matched back to their respective GO Terms for interpretation. The same approach was used for *P. moestus*, only that inverse grep was exclusively used since we had no secondary methacrylic acid producer. Given the smaller set of upregulated genes and thus smaller list of upregulated GO Terms in *Brachinus elongatulus* tissues, we only compared *Goniotropis kuntzeni* and *Pheropsophus jessoensis* for the benzoquinone producers; *B. elongatulus* GO IDs were neither grepped nor inverse grepped as part of this analysis.

Phylogenetic reconstruction of Adephaga

To assess the pattern of defensive chemical evolution in Adephaga, a comprehensive phylogeny suitable for trait mapping is necessary. While such a phylogeny already exists and contains some congeners of our species, we saw several benefits to adding our six

taxa to the existing topology (Vasilikopoulos et al 2021). Our transcriptomes were comparatively more complete compared to the transcriptomic data and genomic data used in previous analyses, making our data potentially useful for resolving uncertain relationships between groups. Also, several of our species did not have congeners in the existing phylogeny, including *Brachinus elongatulus*, *Harpalus pensylvanicus,* and possibly *Goniotropis kuntzeni* given that the *Goniotropis* sp. in the existing topology was designated as *sensu lato* at the genus level. Additionally, by having a robust hypothesis for the evolution of our species, we can draw more informed conclusions about patterns that we may see in gene family phylogenies (e.g. gene discordance, putative lineage-specific gene family expansions/depletions, etc.).

The original article utilized OrthoGraph to extract from several adephagan transcriptomes and genomes sequences that were orthologous to those in a five-species reference set (Petersen et al 2017, Vasilikopoulos et al 2021). We followed the same protocol of utilizing this five-species reference set to run OrthoGraph (v0.7.1) (not installed via miniconda3) on our six transcriptomes, extracting the same putative orthologous from each. A custom script was then used to concatenate all sequences from our six species in an orthogroup-wise manner. For all further steps, we only utilized the 351 orthogroups that constituted the "filtered dataset" in the original paper, discarding all others. As raw sequences were not provided in the published dataset, only the alignments, we used the "--add" flag in MAFFT to align our sequences to the existing alignments (Katoh & Standley 2013). Using TrimAl (v1.4.rev22) (not installed via miniconda3) with the "--gappyout" parameter set, gap-rich regions of each alignment were then removed, including any sequences that were constituted only of gaps (Capella-Gutiérrez et al.

2009). AMAS (v1.0) was used to concatenate the alignments and generate partitions (Borowiec 2016). IQ-Tree (v2.1.4_beta) was run with the same parameters described in the original article for the partitioned datasets (Minh et al 2020). Phylogenies were visualized in R using the ape (v5.5) and ggtree (v3.0.4) packages (Paradis et al. 2004, Yu et al. 2017) (not installed via miniconda3). As the Carabidae were the primary group of interest for us, we subsetted the tree to include only Geadephaga. We chose *Trachypachus gibsii* as our outgroup given its placement as the outgroup to the rest of Geadephaga in the comprehensive phylogeny.

Phylogenetic analysis of candidate gene families involved in defensive chemical biosynthesis and transport

To aid in the functional annotation of candidate genes involved in the biosynthesis and transport of defensive chemicals and to better understand the role convergence may play in the evolution of defensive chemical phenotypes, we reconstructed phylogenies of such candidate gene family members across all six species. These gene families include the sodium:solute transporter family, animal haem peroxidases, and GMC oxidoreductases. Since protein family membership was established during the HMMScan step of functional annotation, a custom script was used to extract gene IDs associated with PFam identifiers from each species' annotation database. The seqtk bioinformatics toolkit was then used to subset the *in silico* translated peptides of each species generating several multifasta files of putative gene family members for each gene family and each species. A custom script was also used to subset these datasets such that only the longest protein per Trinity gene was kept.

Gene family members from all six carabid were then aligned using MAFFT (Katoh & Standley 2013). Default settings were used to generate raw multiple sequence alignments. Summary statistics were then generated using TrimAl and gap-rich regions removed using the --gappyout parameter (Capella-Gutiérrez et al. 2009). If the removal of gap-rich columns in the alignment left any sequence as being solely comprised of gaps, that sequence was automatically removed by TrimAl. Gene family phylogenies were then reconstructed via IQ-Tree using default settings aside from thread usage and specifying 1000 ultrafast bootstrap replicates per gene family (Minh et al 2020). All resulting phylogenies were left unrooted. Given the large size of these phylogenies, clades of interest were subsetted from them and bootstrap support values removed except for critical nodes to aid in visual clarity.

RESULTS

GC-MS Analysis of Pygidial Gland Defensive Chemicals

As described previously, we detected formic acid, undecane, and 2-tridecanone in the pygidial gland sprays of *Harpalus pensylvanicus*, a suite of compounds typical of this genus (Rork et al. 2021). The pygidial glands of *Platynus angustatus* also contained formic acid, undecane, and 2-tridecanone, although we also detected the presence of acetic acid, 2-tetradecanone, and 2-pentadecanone (Fig. 1). To our knowledge, this is the fourth known instance of formic acid being produced by members of the genus *Platynus*, the second instance of 2-tridecanone detection, and the second of undecane detection. The remaining three compounds had not been detected in the pygidial gland defensive secretions of *Platynus* spp. to date (Schildknecht 1970, Kanehisa & Murase 1977,

Kanehisa & Kawazu 1985). The pygidial gland secretions of *Pterostichus moestus* contained five total compounds: acetic acid, propanoic acid, 2-methylpropanoic acid, methacrylic acid, and tiglic acid (Fig. 2). This is the twenty-fourth known instance of methacrylic and tiglic acids in *Pterostichus* spp., the second instance of acetic acid and 2-methylpropanoic acid detection, and the first of propanoic acid detection (Schildknecht 1970, Kanehisa & Murase 1977, Kanehisa & Kawazu 1982, Will et al. 2000).

The pygidial gland secretions of *Brachinus elongatulus* primarily contain 1,4-benzoquinone, 2-methyl-1,4-benzoquinone, 9-docosene, and 9-tricosene as described previously, along with trace amounts of 1,4-hydroquinone, 2-methyl-1,4-hydroquinone, and other trace quinoid compounds (Attygalle et al. 2020). As mentioned, the pygidial gland constituents of *Pheropsophus jessoensis* have been characterized in previous studies (Kanehisa & Murase 1977).

RNA-Seq Statistics

Due to the rather lenient trimming parameters, we lost very little data in terms of total reads across all samples (1.7% to 3.6%, averaged per species). For four species, *Brachinus elongatulus*, *Goniotropis kuntzeni*, *Harpalus pensylvanicus*, and *Pheropsophus jessoensis*, we also lost very little in terms of bases discarded (2.1% to 5.7% averaged per species). For *Platynus angustatus* and *Pterostichus moestus*, unfortunately, we lost substantially more bases due to higher adapter content across these samples (13.2% and 15.1% bases lost, respectively). In these two species, the percent of reads containing adapters were 46.1% and 47.7% respectively, whereas most other species' samples had less than 5% adapter content with a few exceptions in *H. pensylvanicus*. Nevertheless,

due the much higher depth of sequencing for *P. angustatus* and *P. moestus*, these two species still retained a greater number of total number of reads post-trimming than the other four species.

It was likely unnecessary to trim polyG tails given the extremely low percentage of reads trimmed (0.01-2.27%). It also may not have been necessary to trim N-containing reads, which seem to have been similarly rare (0.06-1.93%). In the former case especially, given the rarity of polyG tails despite the sequencer used leveraging two-color chemistry, it may be the case that a significant proportion of those polyG tails trimmed were real guanine homopolymers rather than sequencing artifacts. A full list of summary statistics for the raw and trimmed RNA-Seq data can be found in Table 1.

Transcriptome Assembly Statistics.

All six species' transcriptomes were almost entirely complete according to BUSCO, ranging from 97.4% to 99.5%. In all cases, the majority of BUSCOs present were duplicated (59.0% to 92.8%). Based on the criterion by which BUSCO assesses a gene to be "duplicated" (i.e. more than one copy found in the test set) is probable that this inflated duplicated value is due to the assembly of multiple isoforms per gene rather than true duplication events within the genomes of these beetles (Simão et al. 2015).

In terms of assembly size, *Brachinus elongatulus* had the smallest number of assembled transcripts (49,340) and the shortest total assembly length (82.95 Mb). *Goniotropis kuntzeni* had the largest assembly in terms of both number of assembled transcripts (253,753) and total length (252.41 Mbp). The likely reason for this discrepancy is the relatively low depth of sequencing in *B. elongatulus* compared to other

taxa, or conversely, the high depth of sequencing for *G. kuntzeni* compared to other taxa. GC content was similar for all species, ranging from 36.3% to 38.4%. E90N50 values ranged from 1649 bp to 2096 bp. These values were rather like the unfiltered N50 values, which ranged from 1573 bp to 2581 bp.

In *B. elongatulus* and *Pheropsophus jessoensis*, over half of all transcripts were predicted to have at least one coding sequence (67.9% and 56.0% respectively). In the remaining four species, between 44.7% and 47.7% of all transcripts were predicted to contain at least one coding sequence. Across all six species, between 32.8% and 53.8% of transcripts received one or more blastx hits, between 60.4% and 72.8% of all predicted peptides received one or more blastp hits, and between 53.5% and 66.4% of all predicted peptides received one or more hmmscan hits.

Across all species, the average pseudoalignment rate fell between 68.0% and 97.2%, the lowest being in *Harpalus pensylvanicus* and the highest in *B. elongatulus*. Differential gene expression analyses revealed between 286 and 1010 genes to be significantly upregulated in the secretory lobes of all six species with *B. elongatulus* having the fewest upregulated genes and *H. pensylvanicus* having the most. In the bombardier beetles, *B. elongatulus* had the fewest upregulated genes in the reaction chamber, followed by *G. kuntzeni* with 245 and *P. jessoensis* with 362. A full list of summary statistics for the transcriptome assembly and all downstream analyses can be found in Tables 2-5.

Candidate Genes and Pathways Involved in Formate Biosynthesis

In all subsequent sections, upregulated genes will occasionally be referred to as SL-upregulated or RC-upregulated to reflect upregulation in the secretory lobes relative to the whole body or the reaction chamber relative to the whole body respectively. Note that this specification will be given even for those species without reaction chambers to ensure full clarity.

In the secretory lobes of both formic acid-producers, *Harpalus pensylvanicus* and *Platynus angustatus*, we find evidence for the upregulation of all three genes involved in the core of the folate cycle of C-1 metabolism. Specifically, these are serine hydroxymethyltransferase (*SHMT*), bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase (*MTHFD2*), and trifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, formyltetrahydrofolate synthetase (*MTHFD1*) (Fox & Stover 2008, Brosnan & Brosnan 2016, Meiser et al 2018). In the SL-upregulated gene sets of both *H. pensylvanicus* and *P. angustatus* we find a single copy of *SHMT*. In *H. pensylvanicus,* we find two copies of *MTHFD2* and three copies of *MTHFD1* whereas *P. angustatus* has one copy of *MTHFD2* and two copies of *MTHFD1*. As discussed in previous works, it is unclear whether these multiple copies are real or artifacts of the assembly process (Rork et al. 2021). Not only are these three genes upregulated in the secretory lobes of both formic acid producers, but they are exclusively so; that is, none of the three benzoquinone producers nor the methacrylic acid-producer upregulates any of these genes in their secretory lobes.

We also find evidence for the upregulation of one gene involved in the kynurenine pathway of tryptophan catabolism. Specifically, both species upregulate

kynurenine formamidase (*KFA*), two copies in the SL-upregulated gene set of each (Mehler & Knox 1950, Kanai et al. 2009, Han et al. 2012, Badawy 2017). We find no other genes involved in this pathway to be SL-upregulated in either species' secretory lobes. As with the three folate cycle genes, *KFA* upregulation in the secretory lobes is exclusive to the two formic acid-producers. Although we find evidence for the upregulation of 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase of the methionine salvage pathway in *H. pensylvanicus* secretory lobes, we find no such evidence in *P. angustatus*, nor any evidence for the SL-upregulation of any other gene in the methionine salvage pathway (Savarse et al. 1985, Sekowska et al. 2019).

There are several GO terms enriched exclusively in the secretory lobes of formic acid producers. Amongst those under the ontology "Biological Process" are "Tetrahydrofolate metabolic process", "Folic acid-containing compound metabolic process", and "One-carbon metabolic process", all which map to the core genes of the folate cycle. "Carboxylic acid metabolic process", "Organic acid metabolic process", and "Oxoacid metabolic process" are also exclusively enriched in these species' secretory lobes, all of which again map to genes involved in the folate cycle as well as the kynurenine pathway. Terms exclusively enriched under the "Molecular Function" category are even more specific, listing all primary functions of MTHFD1 and MTHFD2: "Methenyltetrahydrofolate cyclohydrolase activity", "Methylenetetrahydrofolate dehydrogenase (NADP+) activity", and "Formate-tetrahydrofolate ligase activity". Although not specific to KFA, the exclusively enriched term "Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds" does refers to the activity of this enzyme according to the gene IDs mapped to it by GO-Seq.

Candidate Genes and Pathways Involved in Methacrylate Biosynthesis.

In the secretory lobes of *Pterostichus moestus*, we find all genes involved in the valine catabolic pathway through the generation of methacrylyl-CoA to be upregulated (Lange et al. 2004, Wanders et al. 2010). Namely, these are branched-chain-amino-acid aminotransferase *(BCAT)*, the alpha and beta subunits of the E1 subunit of the branched-chain alpha-keto dehydrogenase complex (*BCKDHa* and *BCKDHb*), the E2 subunit of the branched-chain alpha-keto dehydrogenase complex (*DBT*), and a short-chain specific acyl-CoA dehydrogenase (ACAD) (Brosnan & Brosnan 2006, Kochevenko et al. 2012). Only one copy of each gene is present in the upregulated gene set of this species' secretory lobes. These genes are also exclusively upregulated in the secretory lobes of *P. moestus*, our sole methacrylic acid-producer, and not in the secretory lobes of our benzoquinone-producers nor our formic acid-producers.

To our knowledge, there is no known canonical gene implicated in the conversion of methacrylyl-CoA to methacrylic acid. Presumably, such a gene would need to encode an enzyme with thioesterase function or that via some similar mechanism would be able to cleave the CoA molecule from the acyl moiety, hydroxylating the latter to form the corresponding acid (Grevengoed et al. 2014). We do find one SL-upregulated gene likely to have such a function, a probable serine hydrolase. Serine hydrolases comprise a large family of enzymes, many of which function as acyl-CoA hydrolases. Specifically, the serine hydrolase that we observe to be SL-upregulated contains the same α/β hydrolase domain found in other acyl-CoA thioesterases, thus making it likely that this enzyme could hydrolyze methacrylyl-CoA to methacrylic acid and CoA (Grevengoed et al. 2014). Not only that, but given the metabolic cost of synthesizing CoA from vitamin B5, this

enzyme is likely critical to maintaining a high level of free CoA in the secretory lobe cells.

There are also many related GO terms enriched exclusively in the secretory lobes of *P. moestus*. Under “Biological Process” are the closely-related terms "branched-chain amino acid catabolic process" and "branched-chain amino acid metabolic process", both of which describe the breakdown of branch-chained amino acids to their various acyl-CoA derivatives. Under “Cellular Component” are terms relating specifically to the BCKDH complex, namely "dihydrolipoyl dehydrogenase complex", "mitochondrial alpha-ketoglutarate dehydrogenase complex", "tricarboxylic acid cycle enzyme complex", and "mitochondrial tricarboxylic acid cycle enzyme complex".

Candidate Genes and Pathways Involved in Benzoquinone and Hydrogen Peroxide Biosynthesis

In bombardier beetles, the process of benzoquinone biosynthesis occurs in two distinct regions of the pygidial gland system: the secretory lobes and the reaction chamber (Attygalle et al. 2020). The secretory lobes are the site of hydroquinone biosynthesis whereas the reaction chamber is the site of benzoquinone biosynthesis from said hydroquinones.

In *Brachinus elongatulus*, it is known that 2-methyl-1,4-benzoquinone is biosynthesized from 2-methyl-1,4-hydroquinone and that 1,4-benzoquinone is biosynthesized from 1,4-hydroquinone (Attygalle et al. 2020). This process occurs mostly if not entirely within the beetles’ reaction chambers. This step in the pathway has been confirmed through the analysis of reservoir constituents prior to entering the reaction

chamber and spray constituents after leaving the reaction chamber; it has also been confirmed with stable isotope tracing experiments (Attygalle et al. 2020). In the reaction chambers, all three species upregulate animal haem peroxidases which putatively serve to oxidize hydroquinones to their benzoquinone counterparts via the catalytic breakdown of hydrogen peroxide. The latter has been shown to be present in the reservoirs of *Brachinus* sp. at concentrations around 25%, an extraordinarily high concentration by any biological standard (Schildknect 1964). Not only are these genes the most RC-upregulated in terms of log fold change, but they are in all cases the most highly expressed, fully-annotated genes in the reaction chamber.

While the precursor of 2-methyl-1,4-benzoquinone has been demonstrated in brachinines, it has not been demonstrated in paussines. However, given that this appears to be a plausible pathway in Paussinae, we will still reference *Goniotropis kuntzeni* here with the acknowledgement that the true pathway utilized in that lineage may be different.

It is unclear which genes are involved in the conversion of meta-cresol to 2-methyl-1,4-hydroquinone in Brachininae, although it would likely a hydroxylase, oxidase, or oxidoreductase given the relationship between the two compounds. It is worth noting that we do not view it as plausible that the conversion of meta-cresol to 2-methyl-1,4-hydroquinone is accomplished non-enzymatically through hydrogen peroxide-mediated oxidation. While this reaction has been demonstrated to occur, it is typically in the presence of a metal catalyst under laboratory conditions and is rather exothermic. It may also produce more products than just 2-methyl-1,4-hydroquinone such as 4-methyl-1,2-benzoquinone, which has never been detected in Carabidae pygidial gland secretions (Pamin et al. 2015, Morimoto et al. 2015).

No specific hydroxylases are SL-upregulated in either *B. elongatulus* or *Pheropsophus jessoensis*. In *P. jessoensis* and *G. kuntzeni*, we do observe the SL-upregulation of cytochrome P450, many of which have been shown to have oxidoreductase and hydroxylase activity toward aromatic compounds, but to assign them any activity more specific than "oxidoreductase" would be speculative (Schlosser et al. 1993, Gut et al. 1996, Scott & Wen 2001, Yan et al. 2005, Ortix et al. 2010, Schuler & Berenbaum 2013). They have also been implicated in the biosynthesis of various hydrocarbons and hydrocarbon derivatives, and thus may also play a role the biosynthesis of those hydrocarbons found in the pygidial glands (Qiu et al. 2012, Attygalle et al. 2020, Feyereisen 2020). We do not find CYP450s SL-upregulated in the *B. elongatulus* assembly. We also find several laccases upregulated in the secretory lobes of *P. jessoensis*, many of which have confirmed (poly)phenoloxidase activity (e.g. the conversion of simpler phenolics to benzoquinones) (Arakane et al. 2005, Baldrian 2006). This function would be putative in our bombardier beetles. However, they are not SL-upregulated in the *B. elongatulus* or *G. kuntzeni* assembly. Many of these oxidases also tend to convert simpler phenolics all the way to their benzoquinone counterparts, whereas the reservoirs primarily contain hydroquinones with lower levels of benzoquinone presence (Attygalle et al. 2020). We do find it noteworthy that regarding *B. elongatulus*, when examining genes with a slightly lower log fold change cutoff (logFC > 2), both laccases and CYP450s are found to be upregulated in the secretory lobes. Given the comparatively lower power to detect differentially expressed genes in this species than in our other two bombardiers, perhaps it is of some significance that these genes are found to be SL-upregulated with a slightly less stringent filter.

There does appear to be a mechanism by which benzoquinones could be biosynthesized in the secretory lobes from simpler phenolics, only to be reduced to hydroquinones prior to storage in the reservoir. Amongst the most highly expressed and most upregulated genes in all three bombardiers' secretory lobes are members of the GMC oxidoreductase gene family, specifically glucose dehydrogenases. These enzymes play an important role in synthesizing D-glucono-1,5-lactone and hydroquinone from D-glucose and benzoquinone (Ferri et al. 2011, Sygmund et al. 2011). Assuming excess glucose, it may be that bombardier beetles are first converting simpler phenolics to their oxidized benzoquinone counterparts via some oxidase (laccases, CYP450s, etc.), only to reduce them back to their hydroquinone counterparts via glucose dehydrogenase activity. Presumably this process would be highly efficient but not perfect, perhaps leading to the presence of benzoquinones in the reservoir amongst the majority hydroquinones (Attygalle et al. 2020).

Even less is known about the biosynthesis of meta-cresol and 1,4-hydroquinone, including whether they are synthesized in the secretory lobes or the beetle at all. Regarding the latter compound, none of the traditional pathways implicated in hydroquinone biosynthesis in other lineages are likely to be involved in the secretory lobes of the bombardier beetles. We find no evidence for any SL-upregulated genes from the L-tyrosine degradation pathway, the L-phenylalanine degradation pathway, the polyketide synthesis pathway, the decarboxylation of carboxylic aromatic acids, etc. (Meinwald 1966, Blum 1987, van Berkel et al. 1994, Rocha et al. 2013, Raspotnig et al. 2015, Wang et al. 2021). However, we do find several SL-upregulated glycosidases in both *B. elongatulus* and *P. jessoensis*, which may be able to free glycosylated

hydroquinones and cresols from their glycoconjugates (Oba et al. 2013). This is highly theoretical and has not been demonstrated in any Carabidae, but would explain the notable lack of any clear quinone-biosynthesizing or cresol-biosynthesizing genes upregulated in the secretory lobes of any taxa. Were such a scenario supported, that would imply that these two compounds are biosynthesized outside of the secretory lobes, either by the beetles or by endosymbionts, or that they are obtained through the diet (Anbutsu et al. 2017, McManus et al. 2018). Given the capacity of beetles to produce quinones under a range of dietary conditions, we view the lattermost option as less likely.

Regarding the biosynthesis of hydrogen peroxide, we find several upregulated genes in the secretory lobes of both Brachininae with superoxide dismutase annotations (PF00080.21, Sod_Cu). Superoxide dismutases generate hydrogen peroxide from superoxide radicals (O_2^-), which is typically decomposed into water and molecular oxygen by catalases (McCord & Fridovich 1969, Fukai & Ushio-Fukai 2011). However, in bombardier beetles, rather than breaking all the hydrogen peroxide down, they seem to export at least a portion of it to their reservoirs where it is stored until secretion to the reaction chamber. Interestingly, there are no upregulated superoxide dismutases in the secretory lobes of *G. kuntzeni*, suggesting that they may use an alternative means of biosynthesizing hydrogen peroxide. It is also possible that some of our GMC oxidoreductases may truly be glucose oxidases, which have been shown to biosynthesize D-glucono-1,5-lactone and hydrogen peroxide from D-glucose and molecular oxygen (Ferri et al. 2011). Glucose oxidases appear to not have representatives in UniProtKB from metazoans, only from Fungi whose glucose oxidases appear to have low pairwise similarity to insect GMC oxidoreductases (Iida et al. 2007, Sützl et al. 2019). Thus, any

glucose oxidases in our transcriptome may be misannotated as other GMC oxidoreductases or left unannotated altogether. Efforts are underway to assess whether some of our GMC oxidoreductases are in fact glucose oxidases. If so, they would be another plausible means of hydrogen peroxide biosynthesis, including in *G. kuntzeni*.

The GO terms enriched exclusively in the quinone-producers provide little additional clarification regarding which genes are most likely to be involved in hydroquinone biosynthesis in the secretory lobes. Instead, there are several terms enriched that are broadly involved in the "amino-acid betaine metabolic process", the "carbohydrate metabolic process"", and the "choline metabolic process". Genes that map to these terms include laccases, glucose dehydrogenases, choline dehydrogenases and several genes involved in aldose and ketose metabolism such as galactose mutarotase. Also interesting are the terms "pupal chitin-based cuticle development", "chitin-based cuticle development", and "cuticle development" due to the involvement of quinones in insect cuticle development, these mapping almost exclusively to glucose dehydrogenases.

Phylogenetic reconstruction of the Geadephaga

The subfamily-level topology of Geadephaga recovered in our analysis is nearly identical to that of Vasilikopoulos et al. (2021) except for the relationship between Paussinae, Rhysodinae, and Siagoninae (Vasilikopoulos et al. 2021) (Fig. 3). In our analysis, Rhysodinae fell out sister to Paussinae (98% BS) with Siagoninae sister to [Rhysodinae + Paussinae] (100% BS). In Vasilikopoulos et al., Paussinae were sister to Siagoninae with Rhysodinae sister to [Siagoninae + Paussinae]. It is difficult to say which if either topology reflects the true evolutionary history of Rhysodinae, Siagoninae, and Paussinae,

although Vasilikopoulos et al. (2021) did recover this topology as well from the analysis of a subset of their supermatricies.

In any case, all six of our newly added taxa fell out into their appropriate tribes, although two genera became paraphyletic (Bousquet 2012) (Fig. 3). Instead of the two *Platynus* as sister taxa with *Glyptolenus* as the outgroup, our *Platynus angustatus* fell out sister to the original paper's *Platynus* sp. and *Glyptolenus* sp. (100% BS for both nodes) (Fig. 3). Also, the original paper's *Goniotropis sensu lato* sp. fell out as sister to *Ozeana* sp., with our *Goniotropis kuntzeni* as an outgroup rather than the two *Goniotropis* being sister (100% BS for both nodes) (Fig. 3). We view this as a function of data completeness and data processing rather than an issue with specimen identity or cryptic taxonomy. For example, in the supermatrix used the generate the phylogeny in this article, our *P. angustatus* data contained 2.6% gaps. Conversely, *Platynus* sp. and *Glyptolenus* sp. contained 51.7% and 57.4% gaps respectively. This is not unexpected as the datasets were generated for two very different purposes and utilized vastly different depths of sequencing. Several other factors such as our pooling of individuals, difference in RNA extraction, sequencing protocol, bioinformatics approaches, etc., may also play a role. It is important to note that while this phylogeny is the most robust for the family to date in terms of data richness and subfamilies sampled, it misses a large amount of generic diversity in the family, particularly within the Harpalinae. Thus, we cannot confidently assess whether the formic acid chemotype shared by *H. pensylvanicus* and *P. angustatus* is due to convergence or common ancestry, although previous phylogenies that sampled a wider array of taxa with less data do support the notion that is unlikely to be convergent (Ober & Heider 2010, Ober & Maddison 2008). However, it is apparent that our two

benzoquinone producers are rather distantly related, which alongside the observation that quinones can only be found in two additional genera across the phylogeny supports the notion that quinones arose independently in these two lineages (Rork & Renner 2018).

Candidate genes involved in carboxylic acid transport

As important as the ability to synthesize defensive chemicals is to carabids, so too is their ability to secrete them from their secretory lobes to the lumens of the collecting ducts and reservoirs. In our three carboxylic acid producing taxa, this may be accomplished via monocarboxylate symporters, many of which belong to the sodium:solute symporter family (SSF) (Jung 2002). In this section, any reference to an upregulated gene refers to genes upregulated in the secretory lobes.

Phylogenetic reconstruction of the sodium:solute symporter gene family suggests that members of two distinct clades may be important for monocarboxylate transport in the secretory lobes of *Harpalus pensylvanicus*, *Platynus angustatus*, and *Pterostichus moestus* (Fig. 4a,b). *H. pensylvanicus* and *P. angustatus* upregulate members of a clade here denoted as SSF-A (80% BS) whereas *P. angustatus* upregulates members of a clade denotated as SSF-B (93% BS) (Fig. 4a,b). The two upregulated *H. pensylvanicus* genes are sister (100% BS), this clade being sister to a clade comprised of two *P. angustatus* genes (100% BS), which themselves are sister (100% BS), one of which is upregulated. There also appears to be a second *P. angustatus* gene upregulated in SSF-A, although not sister to the aforementioned clade of upregulated genes (Fig. 4a). The one upregulated *P. moestus* sequence, meanwhile, is upregulated in a distantly related clade (SSF-B) with no

other upregulated genes from any of the other six species, although said clade does contain sequences from all other taxa (Fig. 4b).

Interestingly, *Pheropsophus jessoensis* and *Goniotropis kuntzeni* also upregulate sodium:solute symporters, although none being likely orthologues of the *H. pensylvanicus*, *P. angustatus*, or *P. moestus* sequences (Fig. 4a). The upregulation of likely orthologous sodium:solute symporters in the two formic acid producers and a relatively distantly-related sodium:solute symporter in the methacrylic acid producer suggests that if they are involved in the transport of the primary defensive chemicals into the secretory lobe lumens, they are likely transporting different compounds.

Convergence upon GMC oxidoreductase, peroxidase upregulation in bombardier beetles

In this section, any reference to an upregulated GMC gene refers to genes upregulated in the secretory lobes. Phylogenetic reconstruction of the GMC oxidioreductase gene family reveals that upregulated *Brachinus elongatulus*, *Goniotropis kuntzeni*, and *Pheropsophus jessoensis* GMCs cluster together in two clades, aside from one *B. elongatulus* sequence which stands alone in a third clade (Fig 5a,b). These two major clades both have robust support (100% BS each) and will henceforth be referred to as GMC-A and GMC-B. Comparatively, GMC-A is the smaller of the two clades, having one gene from *B. elongatulus*, two genes from *P. jessoensis*, and six genes from *G. kuntzeni* which are upregulated (Fig 5a). GMC-B is larger with two genes from *B. elongatulus*, nineteen genes from *P. jessoensis*, and seven genes from *G. kuntzeni* (Fig. 5b). These numbers may not reflect the true number of genes in these species which are upregulated in the

secretory lobes, but *de novo* annotated genes which may be inflated for species sequenced deeply (i.e. *P. jessoensis* and *G. kuntzeni*) and an undercount for species with low sequencing depth (i.e. *B. elongatulus*).

GMC-A can be further broken down into subclades GMC-A1 (100% BS) and GMC-A2 (100% BS) (Fig. 5a). GMC-A1 is comprised of both *P. jessoensis* and *G. kuntzeni* sequences, but only the latter are upregulated. GMC-A2 is the inverse of this, where all three species' sequences are present, but only the *B. elongatulus* and *P. jessoensis* sequences are upregulated. An identical topology is true for GMC-B (Fig. 5b). Two subclades of GMC-B are recovered with strong support, GMC-B1 (100% BS) and GMC-B2 (100% BS). GMC-B1 is comprised of both *P. jessoensis* and *G. kuntzeni* sequences, but only the latter are upregulated. GMC-B2 is the inverse of this, where all three species' sequences are present, but only the *B. elongatulus* and *P. jessoensis* sequences are upregulated. No sequences of *H. pensylvanicus*, *P. angustatus*, or *P. moestus* are found in GMC-A1, GMC-A2, GMC-B1, or GMC-B2. However, within the larger GMC-A and GMC-B clades there are two small subclades robustly sister to GMC-A2 (97% BS) and GMC-B2 (100% BS) that are comprised entirely of non-upregulated sequences from the three harpalines and no bombardiers (GMC-AH and GMC-BH).

All three bombardier species also upregulate animal haem peroxidases exclusively in their reaction chambers; thus, here any reference to an upregulated animal haem peroxidase gene refers to genes upregulated in the reaction chambers. Phylogenetic reconstruction of the animal haem peroxidase phylogeny demonstrates a single clade (100% BS) containing all upregulated animal haem peroxidases from the reaction chambers of these species (Fig. 6). Within this larger clade are two subclades, which we

denote here as PERO-A and PERO-B (99% BS each). Both clades contain sequences from Brachininae and Paussinae, but PERO-A only contains upregulated sequences from the two brachinines whereas PERO-B only contains upregulated sequences from the paussine.

It is difficult to assess the degree to which any of these upregulated genes from their respective clades are orthologous or paralogous to members of the same clade, or if they are even properly *de novo* annotated as belonging to distinct genomic loci given the lack of genomes to confirm loci of origin. However, the sequences comprising the each of two major GMC glades and the main peroxidase clade are likely closely related to members of their clades based on ultrafast bootstrap support values and likely share similar if not identical functions. It is also rather likely based on the large number of bombardier beetle sequences in these clades relative to non-bombardier beetle sequences that the GMC oxidoreductases and peroxidases have undergone expansions in the Paussinae and Brachininae. Alternatively, these gene families may have expanded early in the Carabidae with Harpalinae having experienced a loss of them. Also interesting is that rather than following the topology of the species tree, bombardier beetle GMCs and bombardier beetle peroxidases cluster together rather than Paussinae sequences being sister to clades of Brachininae and Harpalinae sequences. While this could be due to a loss of genes within the Harpalinae, it also raises the possibility that Paussinae and Brachininae genes have converged on similar motifs at the amino acid level.

DISCUSSION

Defensive-grade formic acid likely biosynthesized similarly in *Harpalus* and *Platynus*

The results of our differential gene expression analyses and GO enrichment analyses suggest up to two pathways may be involved in formic acid biosynthesis in both *Harpalus pensylvanicus* and *Platynus angustatus*. The primary candidate pathway thought to be involved in defensive-grade formic acid biosynthesis is the folate cycle of one-carbon metabolism, which has also been implicated in formic acid biosynthesis in ants (Hefetz & Blum 1978a, Hefetz & Blum 1978b, Rork et al. 2021). The core of this cycle is comprised of three enzymes: serine hydroxymethyltransferase, bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, and trifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, formyltetrahydrofolate synthetase (Fox & Stover 2008, MacFarlane et al. 2009, Brosnan & Brosnan 2016). While tetrahydrofolate is a necessary input to the pathway, the one-carbon donor and precursor to formic acid itself is often L-serine (Cβ) (Hefetz & Blum 1978a, Hefetz & Blum 1978b). This pathway is also reversible, capable of generating L-serine from formate and tetrahydrofolate. The stable intermediates in this pathway, 5,10-methylenetetrahydrofolate (5,10-mTHF) and 10-formyletrahydrofolate (10-fTHF), are utilized for the biosynthesis of thymidylate and purines respectively (Hartman & Buchanan 1959, Carreras & Santi 1995, Ben-Sahra et al. 2016). When the former is reduced by methylenetetrahydrofolate reductase (MTHFR), the 5-methyltetrahydrofolate (5-mTHF) product is also utilized to regenerate methionine from homocysteine via methionine synthase (Zheng et al. 2019). Excess formate is typically exported from the

cell into the extracellular space or recycled by the folate cycle given its cytotoxicity (Meiser et al. 2016).

The second candidate pathway potentially involved in defensive-grade formic acid biosynthesis is the kynurenine pathway. This pathway begins by the oxidation of the imidazole ring of tryptophan, generating N'formylkynurenine in the process (Mehler & Knox 1950, Han et al. 2012, Badawy 2017). N'formylkynurenine is then hydrolyzed by kynurenine formamidase, cleaving off the formyl group as formate and generating L-kynurenine in the process. L-kynurenine can then be used to synthesize a variety of compounds, most notably nicotinamide adenine dinucleotide (NAD^+) (Badawy 2017).

Of these two, we view it more likely that the folate cycle is the primary pathway responsible for defensive-grade formic acid biosynthesis for two primary reasons. The first is that L-serine is much less bioenergetically expensive than L-tryptophan, making it a more suitable precursor for the biosynthesis of relatively large quantities of formate (Moura et al. 2013, Barik 2020, Wu et al. 2020). Not only is L-tryptophan typically the least abundant amino acid in organisms, but animals cannot biosynthesize it (Moura et al. 2013). L-serine, on the other hand, is comparatively more abundant in organisms and can be biosynthesized by animals, typically from L-glycine or 3-phosphoglycerate (Moura et al. 2013, Wu et al. 2020). The second reason relates to the relative metabolic stress likely imposed by the upregulation of each pathway. By nature of being a cycle, the only outputs of the folate cycle when flowing in the direction of formate biosynthesis are formate, L-glycine, NAD(P)H, and ATP (Fox & Stover 2008). With formate being exported to the extracellular space, this leaves only an accumulation of L-glycine, NAD(P)H, and ATP, the latter two of which are likely needed in abundance due to the

high metabolic activity of the glands. The kynurenine pathway on the other hand is unidirectional and acyclic, and thus would be producing a large quantity of comparatively specific metabolites that would need to serve some metabolic purpose (Badawy 2017). Of course, it is possible that the products of the kynurenine pathway are necessary for secretory lobe function, but combined with the bioenergetic cost of using L-tryptophan as the primary precursor to formate biosynthesis, we view this as the less parsimonious of the two options.

In our previous work, we also suggested the methionine salvage cycle as a potential albeit unlikely pathway for formate biosynthesis in *H. pensylvanicus*, due to its SL-upregulation of 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (Rork et al. 2021). While this gene is still SL-upregulated in the new *H. pensylvanicus* assembly, it is not in *P. angustatus*. While that alone is not reason enough to suggest it is unlikely to be significantly involved in formate biosynthesis, the methionine salvage cycle shares many of the same conceptual issues at the kynurenine pathway and thus is not likely to be a major contributor in our view. It is also possible that it does have some limited involvement in formate biosynthesis in *H. pensylvanicus*, its role in *P. angustatus* perhaps just being even more limited if there is one.

It is important to note that while both *H. pensylvanicus* and *P. angustatus,* members of the same subfamily, seem to SL-upregulate the same genes for formic acid biosynthesis, they likely did not converge upon this phenotype. Although our phylogeny is not comprehensive enough at the tribal level to assess patterns of formic acid evolution, prior work supports the notion that the formic acid chemotype was derived from a common ancestor in these two genera (Ober & Heider 2010, Ober & Maddison

2008). However, our data nevertheless can be seen as two independent data points supporting the hypothesis that the folate cycle and perhaps the kynurenine pathway are involved in defensive-grade formic acid biosynthesis in these taxa.

The valine catabolic pathway mostly explains methacrylic acid biosynthesis in *Pterostichus*

The results of our differential gene expression analyses and GO enrichment analyses in *Pterostichus moestus* suggest the involvement of the valine catabolic pathway in the biosynthesis of defensive-grade methacrylic acid. This pathway begins with the deamination of L-valine to form 3-methyl-2-oxobutanoate, which is then decarboxylated and covalently bound to coenzyme A to form isobutyryl-CoA (Brosnan & Brosnan 2006, Wanders et al. 2010). Isobutyryl-CoA is then reduced to methacrylyl-CoA, which can be further catabolized to form succinyl-CoA or propanoyl-CoA, the former being an intermediate in the citric acid cycle and the latter being involved in a variety of metabolic pathways (Wanders et al. 2010). Necessarily, some enzyme capable of hydrolyzing the thioester bond of methacrylyl-CoA would then be capable of generating the final methacrylic acid product. We hypothesize that this enzyme is a serine hydrolase, many of which have been functionally characterized as having acyl-CoA ligase or thioesterase activity (Long & Cravatt 2011, Grevengoed et al. 2014). The serine hydrolase upregulated in the secretory lobes of *P. moestus* specifically belongs to the α/β-hydrolase family which, while a poorly characterized group overall, does contain members with esterase activity possibly capable of hydrolyzing methacrylyl-CoA to methacrylic acid

(Long & Cravatt 2011). This of course is speculative and would require functional characterization to confirm.

Interestingly, several other compounds present in the pygidial secretions of *P. moestus* are likely generated through the same metabolic module, the wider catabolism of branched-chain amino acids. L-valine, L-leucine, and L-isoleucine can all be reduced to isobutyryl-CoA, isovaleryl-CoA, and (S)-2-methylbutanoyl-CoA by the activity of BCAT and the E1/E2 subunits of the BCKDH complex respectively (Wanders et al. 2010, Brosnan & BRosnan 2006, Kochevenko et al. 2012). Through the activity of generalist short-chain acyl-CoA dehydrogenases or substrate-specific dehydrogenases, these molecules are further reduced to methacrylyl-CoA, 3-methylcrotonyl-CoA, and tiglyl-CoA respectively. Isobutyryl-CoA, methacrylyl-CoA, and tiglyl-CoA are all simply the acyl-CoA esters of the carboxylic acids 2-methylpropanoic acid, methacrylic acid, and tiglic acid respectively. Throughout the higher-order carboxylic acid-defended Carabidae, several other carboxylic acids such as crotonic acid, isobutyric acid, isovaleric acid, angelic acid, etc., are likely also produced via these three interconnected pathways (Moore & Wallbank 1968, Kanehisa & Murase 1977). The relationship between these pathways and the defensive chemistry of such Carabidae is further supported by evidence that the stable isotope D8-L-Valine is incorporated into methacrylic and isobutyric acids in *Scarites subterraneus*, that [2,3,4,4-(2)H(4)]isoleucine is incorporated into tiglic, 2-methylbutyric, and ethacrylic acids in *Pterostichus californicus*, and that L-[U-14C]-Valine, L-[U-14C]-Leucine, and L-[U-14C]-Isoleucine were incorporated into methacrylic acid, 3-methylcrotonic acid, and tiglic acid respectively in *Carabus yaconinus* (Adachi et al. 1985).

Benzoquinone biosynthesis involves the oxidation of simpler phenolics

Several pathways for the biosynthesis of cresols and quinones have been hypothesized and experimentally verified across life, including in insects. The Tenebrionidae have been shown to biosynthesize p-benzoquinones from L-tyrosine and L-phenylalanine, as well as acetate via the polyketide pathway (Meinwald et al. 1966, Blum 1987). The latter is also true of Gonyleptidae harvestmen which biosynthesize p-benzoquinones from acetate and propionate (Rocha et al. 2013). The fungus *Aspergillus nidulans*, which biosynthesizes 1,4-benzoquinone from a meta-cresol precursor, synthesizes the latter via the decarboxylation of 6-methylsalicylic acid (Wang et al. 2021). In an analogous pathway, *Candida parapsilosis* biosynthesizes hydroquinone directly from 4-hydroxybenzoate, here too via oxidative decarboxylation (van Berkel et al. 1994). Other lineages appear to not necessarily biosynthesize hydroquinones themselves, but to store them in a glycosylated form such as arbutin, as is the case with the firefly *Luciola lateralis* and as is hypothesized to be the case in the termite *Mastotermes darwiniensis* (Oba et al. 2013, He et al. 2018).

Previous work has shown that *Brachinus elongatulus* biosynthesizes 2-methyl-1,4-benzoquinone from 2-methyl-1,4-hydroquinone, and 1,4-benzoquinone from 1,4-hydroquinone in the reaction chamber (Attygalle et al. 2020). It has been established that peroxidases play a role in this process via one-electron oxidation using hydrogen peroxide as an electron donor (Schildknect 1964). In the reaction chambers of all three bombardier beetle species, not only are peroxidases the most upregulated genes in terms of log fold change, but they are the most highly expressed, fully-annotated genes.

Further, in *B. elongatulus* it has been shown that 2-methyl-1,4-hydroquinone is biosynthesized from meta-cresol via an oxidation or hydroxylation reaction (Attygalle et al. 2020). Cytochrome P450s, many of which have been characterized to have phenoloxidase activity, may play a role in this reaction, although their upregulation is not found in all bombardier beetle secretory lobes using a logFC > 4 cutoff and they tend to convert phenolics to benzoquinones (Schlosser et al. 1993, Gut et al. 1996, Scott & Wen 2001, Schuler & Berenbaum 2013). Alternatives include laccases, phenoloxidases part of the multicopper oxidase family, although these also tend to oxidize phenolics to benzoquinones (Arakane et al. 2005). Like cytochrome P450s, they too are also not found in all bombardier beetle secretory lobes at logFC > 4. However, in all three bombardiers, there is at least one upregulated gene in the secretory lobes with logFC > 2 capable of (phenol)oxidase activity, even if not all three species share the same set of upregulated (phenol)oxidase genes. In any case, this still raises the question: "What role could enzymes that primarily convert phenolics fully to benzoquinones possibly play in hydroquinone biosynthesis?"

We propose a solution to the apparent lack of enzymes that can convert meta-cresol directly to 2-methyl-1,4-hydroquinone without converting it all the way to benzoquinone: glucose dehydrogenases of the GMC oxidoreductase gene family. These enzymes have the capacity in the presence of excess glucose to convert benzoquinones to hydroquinones, which would allow for meta-cresol to be oxidized directly to 2-methyl-1,4-benzoquinone by phenoloxidases such as laccases or CYP450s (Ferri et al. 2011). This is not inconsistent with incorporations results, as meta-cresol would still ultimately become the 2-methyl-1,4-benzoquinone we see in bombardier beetle sprays, it would just

take that form twice: once prior to leaving the secretory lobes for the reservoir and once after leaving the reaction chamber. This may also explain the presence of benzoquinones in quinone-producers' reservoirs, which is often explained as being due to contamination from the reaction chamber (Attygalle et al. 2020). Thus, rather than the biosynthetic scheme being "meta-cresol (SL) to 2-methyl-1,4-hydroquinone (SL) to 2-methyl-1,4-benzoquinone (RC)", we propose that a more likely pathway may be "meta-cresol (SL) to 2-methyl-1,4-benzoquinone (SL) to 2-methyl-1,4-hydroquinone (SL) and back to 2-methyl-1,4-benzoquinone (RC)" (SL=secretory lobes, RC=reaction chamber).

Regarding any difference between *Pheropsophus jessoensis* and *Goniotropis kuntzeni*, this could simply be due to the evolution of divergent strategies for arriving at similar chemotypes (i.e. benzoquinones biosynthesized in different ways). In fact, this may be expected given that there is a plethora of enzymes capable of catalyzing the conversion of simpler phenolics to quinone counterparts, even more-so if these pathways are novel and aren't necessarily constrained by highly specific reaction mechanisms.

We find no evidence for any genes that are likely to be involved in the biosynthesis of 1,4-hydroquinone, 2-methyl-1,4-hydroquinone, or meta-cresol via L-tyrosine catabolism, L-phenylalanine catabolism, the polyketide pathway, decarboxylation of aromatic carboxylates, or any other putative pathway. We do find several glycosidases upregulated in the secretory lobes of all bombardier beetle taxa, which may play a role in generating meta-cresol or 1,4-hydroquinone from a glycoside, but this wouldn't necessarily explain the biosynthesis of the aromatic ring itself (Oba et al. 2013, He et al. 2018). Given the ability of bombardiers to produce benzoquinones whether fed on mealworms, dog food, fruit, etc., and whether fed or injected with

labelled precursors, it is unlikely that a specific dietary source of arbutin or phenolic glycosides is necessary (Attygalle et al. 2020). It is also possible that endosymbionts may play a role in generating 1,4-hydroquinone, meta-cresol, or precursors further upstream in the pathway (Anbutsu et al. 2017, McManus et al. 2018). This would potentially explain the lack of any clear candidate genes involved in meta-cresol or 1,4-hydroquinone biosynthesis, but also lacks support. It is also quite possible that there is evidence of genes involved in the biosynthesis of meta-cresol and 1,4-hydroquinone in the secretory lobes, but that they are uncharacterized (*de novo* biosynthesis) or have been annotated in such a way that their role in such a process is not obvious to us.

Importantly, due to the observation that all members of Brachininae are quinone-producing bombardier beetles, similarities in the transcriptomes or gene expression profiles of *B. elongatulus* and *P. jessoenis* are likely due to shared common ancestry. However, the Paussinae have been shown to be an independent lineage of bombardier beetles with robust support. Similarities in the transcriptomes or gene expression profiles of Brachininae and Paussinae, especially as they pertain to benzoquinone biosynthesis, are more likely the results of convergent or parallel evolution than shared common ancestry. It is also possible that differences between *B. elongatulus* and *P. jessoensis* may reflect true differences in glandular gene regulation or may be a function of differing power to detect differentially expressed genes between the two species.

Convergence upon pygidial gland metabolism in bombardier beetles

GMC oxidoreductases in the secretory lobes and animal haem peroxidases in the reaction chamber are the two likely candidates to be involved in benzoquinone metabolism based

on our present knowledge. The former likely play little role in the biosynthesis of 1,4-hydroquinone or 2-methyl-1,4-hydroquinone from comparatively reduced precursors, but they may be involved in converting benzoquinones to hydroquinones in the secretory lobes as discussed previously (Ferri et al. 2011, Sygmund et al. 2011). The latter are well-established to play a role in the catalytic decomposition of hydrogen peroxide in the reaction chamber, allowing for the characteristic blasting defensive phenotype responsible for the bombardier beetles' namesake (Hurd et al. 2015).

Phylogenetic analysis of the GMC oxidoreductase family suggests that all three bombardier beetles upregulate two major clades on GMC oxidoreductases in their secretory lobes, here referred to as GMC-A and GMC-B. Both clades share similar topologies, having two bombardier beetle-specific subclades (GMC-A1, GMC-A2, GMC-B1, and GMC-B2), as well as one harpaline-specific subclade (GMC-AH and GMC-BH). Relative to our own topology of Carabidae and previously published topologies, these clades are discordant. This suggests that either the Harpalinae have lost a large suite of genes present in the rest of Carabidae including Brachininae and Paussinae, or that Brachininae and Paussinae have experienced expansions in these gene families with some genes converging upon specific motifs at the amino acid level. We view the latter scenario as more likely. It does seem plausible that certain amino acid motifs may make peroxidases more efficient at decomposing peroxides and GMCs more efficient at maintaining high hydroquinone to benzoquinone ratios. Such motifs would likely be favored by natural selection, especially given the very similar biochemical makeups of these beetles' glandular secretions. Also, while it is possible that a tandem gene cluster could be lost in a common ancestor of Harpalinae, it seems less likely that

they would lose one for GMCs and one for peroxidases separately. In any case, regarding the possibility of gene family expansions, it is important to note that these are *de novo* transcriptomes. Thus, there is little guarantee that even with the relatively large number of genes annotated for the bombardiers that these represent distinct loci in the genomes. Further genomic work will likely be needed to assess to degree to which these gene families have expanded in bombardiers or have experienced losses in Harpalinae.

Formic and methacrylic acid-producers may export their defensive chemicals differently

Although it is ultimately uncertain which genes are involved in defensive chemical export from the secretory lobes in all taxa, we see the most likely candidates in the carboxylic acid producers as belonging to the sodium:solute symporter family. Specifically, this family includes proteins involved in monocarboxylate transport, which as the name implies include compounds such as branched-chain fatty acids and formic acid (Jung 2002, Ganapathy et al. 2008, Moschen et al. 2012). Indeed, these acids have been shown to be transported by members of this gene family in other systems.

Although the exact functions of these transporters remain unclear without experimental validation, it is possible given that they fall out into two rather distantly related clades that they have different substrate specificities. One may be more suitable to transporting the comparatively small, polar, and acidic formic acid while one may be more suitable to transporting the larger, less polar, less acidic methacrylic acid. It is also possible that neither of these transporters are implicated in the export of defensive chemicals, but are rather involved in the transport of important metabolites such as

pyruvate, lactate, etc. Alternatives to monocarboxylate transporters that may be involved in the transport of formic and methacrylic acids across cell membranes include a variety of organic anion and cation transporters, many of which are members of the functionally diverse major facilitator superfamily (MFS) (Quistgaard et al. 2016). However, those few upregulated in the secretory lobes of *Platynus angustatus* generally seem unlikely to be involved in formate transport, but rather trehalose transport. While this does not exclude the possibility that *Harpalus pensylvanicus* or *Pterostichus moestus* may be using MFS or other gene family members instead of or in conjunction with SSF gene family members for defensive chemical export, it seems to reduce the likelihood that *P. angustatus* are.

CONCLUSIONS

The Carabidae are a taxonomically and biochemically diverse lineage whose chemical defense strategies have fascinated evolutionary biologists since the time of Darwin. Over the past several decades, considerable efforts have been undertaken to characterize the diversity of these defensive chemicals, leading to the accrual of hundreds of known compounds from carboxylic acids to phenolics, from sulfides to cyanide and more. Considerable interest over the years has been generated regarding how these beetles (and insects more broadly) synthesize their defensive compounds, but little evidence has been accumulated in only a handful of taxa. Here, we provide transcriptomic evidence suggesting likely pathways for the biosynthesis of the two most common defensive chemicals across the Carabidae, formic acid and methacrylic acid. This work builds upon prior research suggesting that defensive-grade formic acid is biosynthesized by insects from serine via the folate cycle and that defensive-grade methacrylic acid is

biosynthesized in Carabidae through the catabolism of valine. We also provide evidence for a pathway not previously suggested to play a major in formic acid biosynthesis in insects, the kynurenine pathway. While the biosynthetic pathway leading to benzoquinone biosynthesis in carabid remains unclear, we provide convincing transcriptomic evidence for the role that peroxidases play in converting hydroquinones to benzoquinones and narrow down the likely list of candidate genes and pathways involved in the larger process.

Future directions in our view should focus on two primary areas of inquiry: 1) increasing the volume of glandular transcriptomics data available for the Carabidae, focusing both on achieving greater taxonomic and chemical breadth, and 2) verifying the roles of candidate pathways mentioned herein via stable isotope tracing and RNAi-mediated gene knockdown of candidate genes. By increasing the amount of transcriptomics data gathered, we enhance both our capacity to narrow down candidate genes and pathways and study instances of convergence and parallel evolution in lineages where defensive chemicals appear to have arisen independently. Functional verification would be most useful in elucidating biosynthetic processes yet shrouded in uncertainty, such as those of the benzoquinones and the last step of methacrylic acid biosynthesis from methacrylyl-CoA. However, it is still important in our view to functionally validate even the relatively likely candidate pathways such that alternative hypotheses can be ruled out.

ACKNOWLEDGEMENTS

We would like to thank Yuka Imamura and Teodora Orendovici for their assistance in both preparing and sequencing our cDNA libraries at The Pennsylvania State University. We owe great thanks to Athula Attygalle and Sihang Xu for facilitating the GC-MS analysis of our carabids' defensive chemicals, and to Wendy Moore and Kip Will for providing us with the bombardier beetles' RNA-Seq data, as well as the materials and methods used in their collection and processing. We would also like to thank Jong Kyun Park for his assistance collecting *Pheropsophus jessoensis* in South Korea. We thank Julie Urban, Etya Amsalem, Shaun Mahony, Kylie Bocklund, Chloe P. Drummond, Kadeem Gilbert, Blue Riggins, Yunjia Liu, Sean Perez, and Reilly McManus for their helpful discussion on technical aspects of this research as well as the composition of this manuscript. This work was supported by the National Science Foundation [DEB 1556898 (Athula Attygalle), 1556813 (Wendy Moore), 1762760 (Tanya Renner), and 1556957 (Kipling Will)] and The Coleopterists Society Graduate Student Research Enhancement Award. The findings and conclusions of this work do not necessarily reflect the views of the National Science Foundation or The Coleopterists Society.

REFERENCES

1. Adachi Y, Kanehisa K, Tsumuki H. (1985). Catabolic formation of methacrylic acid and tiglic acid from the branched chain amino acids in the pygidial defensive glands of *Carabus yaconinus* Bates (Coleoptera: Carabidae). Appl Entomol. Zool. 20: 492–493.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990). Basic local alignment search tool. Journal of Molecular Biology. 215: 403-410.
3. Anbutsu H, Moriyama M, Nikoh N, Hosokawa T, Futahashi R, Tanahashi M, Meng X, Kuriwada T, Mori N, Oshima K, Hattori M, Fujie M, Satoh N, Maeda T, Shigenobu S, Koga R, Fukatsu R. (2017). Small genome symbiont underlies cuticle hardness in beetles. Proceedings of the National Academy of Sciences of the United States of America. 114(40): E8382-E8391.
4. Aneshansley DJ, Eisner T, Widom JM, and Widom B. (1969). Biochemistry at 100 °C: Explosive Secretory Discharge of Bombardier Beetles (Brachinus). Science. 165(3888): 61-63.
5. Arakane Y, Muthukrishnan S, Beeman RW, Kanost MR, Kramer KJ. (2005). Laccase 2 is the phenoloxidase gene required for beetle cuticle tanning. Proceedings of the National Academy of Sciences of the United States of America. 102(32): 11337-11342.
6. Arndt E, Beutel RG, and 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). (pp. 119–144). Walter de Gruyter, Berlin

7. Attygalle AB, Meinwald J, Eisner T. (1991). Biosynthesis of methacrylic and isobutyric acids in a carabid beetle, *Scarites subterraneus*. Tetrahedron Letters. 32: 4489–4852.
8. Attygalle AB, Morgan ED. (1985). Ant Trail Pheromones. Advances in Insect Physiology. 18: 1-30.
9. Attygalle AB, Smedley SR, Meinwald J, Eisner T. (1993). Defensive secretion of two notodontid caterpillars (*Schizura unicornis*, *S. badia*). Journal of Chemical Ecology. 19(10): 2089-2104
10. Attygalle AB, Wu X, and Will KW. (2006). Biosynthesis of Tiglic, Ethacrylic, and 2-Methylbutyric Acids in a Carabid Beetle, *Pterostichus* (*Hypherpes*) *californicus*. Journal of Chemical Ecology. 33(5): 963-970.
11. Attygalle AB, Wu X, Maddison DR, Will KW. (2009). Orange/lemon-scented beetles: opposite enantiomers of limonene as major constituents in the defensive secretion of related carabids. Naturwissenschaften. 96: 1443-1449.
12. 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*. The Science of Nature 107: 26.
13. Badawy AA. (2017). Kynurenine Pathway of Tryptophan Metabolism: Regulatory and Functional Aspects. International Journal of Tryptophan Research. 10: 1-20.
14. Baldrian P. (2006). Fungal laccases – occurrence and properties. FEMS Microbiology Reviews. 30(2): 215-242.

15. Barik S. (2020). The Uniqueness of Tryptophan in Biology: Properties, Metabolism, Interactions and Localization in Proteins. International Journal of Molecular Sciences. 21(22): 8776.
16. 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
17. Benjamini Y, Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. 57(1): 289–300
18. Birch MC, Poppy GM, Baker TC. (1990). Scents and Eversible Scent Structures of Male Moths. Annual Review of Entomology. 35: 25-58
19. Blum MS. (1987). Biosynthesis of arthropod exocrine compounds. Annual Review of Entomology. 32: 381–413
20. Borowiec ML. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 4: e1660.
21. Bousquet Y. (2012). Catalogue of Geadephaga (Coleoptera, Adephaga) of America, north of Mexico. Zookeys. 245: 1-1722.
22. Bray NL, Pimentel H, Melsted P, Pachter L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology. 34: 525-527.
23. Brosnan JT, Brosnan ME. (2006). Branched-chain amino acids: enzyme and substrate regulation. The Journal of Nutrition. 136(1): 207S-211S.
24. Brosnan ME & Brosnan JT. (2016). Formate: The neglected member of one-carbon metabolism. Annual Review of Nutrition. 36: 369-388.

25. Bushnell B. (2014). BBTools software package.

26. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 25(15): 1972-1973.

27. Carreras CW, Santi DV. (1995). The catalytic mechanism and structure of thymidylate synthase. Annual Review of Biochemistry. 64: 721-762.

28. Crowson RA. (1960) The phylogeny of Coleoptera. Annual Review of Entomology. 5: 111–134.

29. Davidson NM, Hawkins ADK, Oshlack A. (2017) SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes. Genome Biology. 18: 148.

30. Deuve T. (1993). L'abdomen et les genitalia des femelles de Coleopteres Adephaga. Memoires du Museum National D'Histoire Naturelle. 155: 1-184.

31. Di Giulio A, Muzzi M, Romani R. (2015). Functional anatomy of the explosive defensive system of bombardier beetles (Coleoptera, Carabidae, Brachininae). Arthropod Structure and Development. 44(5): 468-490.

32. Eddy SR. (1998). Profile hidden Markov models. Bioinformatics. 14(9): 755-763.

33. 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). Journal of Insect Physiology. 6(4): 272-292.

34. Ferri S, Kojima K, Sode K. (2011). Review of glucose oxidases and glucose dehydrogenases: a bird's eye view of glucose sensing enzymes. Journal of Diabetes Science and Technology. 5(5): 1068–1076.

35. Feyereisen R. (2020). Origin and evolution of the CYP4G subfamily in insects, cytochrome P450 enzymes involved in cuticular hydrocarbon synthesis. Molecular Phylogenetics and Evolution. 143: 106695.
36. Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Research. 39(Web Server issue): W29-W37.
37. Forsyth DJ. (1972). The structure of the pygidial defence glands of Carabidae (Coleoptera). Transactions of the Zoological Society of London. 32: 249-309.
38. Forsyth, DJ. (1968). The structure of the defence glands in the Dytiscidae, Noteridae, Haliplidae and Gyrinidae (Coleoptera). Transactions of the Royal Entomological Society of London. 120: 158-181.
39. Forsyth, DJ. (1970). The structure of the defence glands of the Cicindelidae, Amphizoidae, and Hygrobiidae (Insecta: Coleoptera). Journal of Zoology, London. 160: 51-69.
40. Fox JT, Stover PJ. (2008). Folate-mediated one-carbon metabolism. Vitamins and Hormones. 79: 1-44.
41. Fukai T, Ushio-Fukai M. (2011). Superoxide dismutases: role in redox signaling, vascular function, and diseases. Antioxidants & Redox Signaling. 15(6). 1583–1606.
42. Ganapathy V, Thangaraju M, Gopal E, Martin PM, Itagaki S, Miyauchi S, Prasad P. D. (2008). Sodium-coupled monocarboxylate transporters in normal tissues and in cancer. The AAPS Journal. 10(1): 193–199.
43. 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. Nature Biotechnology. 29(7): 644-652.

44. Grevengoed TJ, Klett EL, Coleman RA. (2014). Acyl-CoA metabolism and partitioning. Annual Review of Nutrition. 34: 1-30.
45. Gustafson GT, Alexander A, Sproul JS, Pflug JM, Maddison DR, Short AEZ. (2019). Ultraconserved element (UCE) probeset design: base genome and initial design parameters critical foroptimization. Ecology and Evolution. 9: 6933–6948.
46. Gustafson GT, Baca SM, Alexander AM, Short, AEZ (2020). Phylogenomic analysis of the beetle suborder Adephaga with com-parison of tailored and generalized ultraconserved element probe per-formance. Systematic Entomology. 45: 552–570.
47. Gut I, Nedelcheva V, Soucek P, Stopka P, Tichavská B. (1996). Cytochromes P450 in benzene metabolism and involvement of their metabolites and reactive oxygen species in toxicity. Environmental Health Perspectives. 104 Suppl 6(Suppl 6): 1211–1218.
48. 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. Nature Protocols. 8(8): 1494-1512.
49. Han Q, Robinson H, Li J. (2012). Biochemical identification and crystal structure of kynurenine formamidase from *Drosophila melanogaster*. Biochemical Journal. 446(2): 253-260.

50. Hartman SC, Buchanan JM (1959). Biosynthesis of the purines. XXVI. The identification of the formyl donors of the transformylation reactions. The Journal of Biological Chemistry. 234(7): 1812–1816

51. He S, Johnston PR, Kuropka B, Lokatis S, Weise C, Plarre R, Kunte HJ, McMahon. (2018. Termite soldiers contribute to social immunity by synthesizing potent oral secretions. Insect Molecular Biology. 27(5): 564-576.

52. Hefetz A, Blum MS. (1978a). Biosynthesis and accumulation of formic acid in the poison gland of the carpenter ant *Camponotus pensylvanicus*. Science. 201(4354): 454-455.

53. Hefetz A, Blum MS. (1978b). Biosynthesis of formic acid by the poison glands of formicine ants. Biochimica et Biophysica Acta. 543(4): 484-496

54. Hurd TR, Liang FX, Lehmann R. (2015). Curly Encodes Dual Oxidase, Which Acts with Heme Peroxidase Curly Su to Shape the Adult *Drosophila* Wing. PLOS Genetics. 11(11): e1005625.

55. Iida K, Cox-Foster DL, Yang X, Ko WY, Cavener DR. (2007). Expansion and evolution of insect GMC oxidoreductases. BMC Evolutionary Biology. 7: 75.

56. Jung H. (2002). The sodium/substrate symporter family: structural and functional features. FEBS Letters. 1(2): 73-77.

57. 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. Molecular Brains. 2: 8.

58. Kanehisa K, Kawazu K. (1982). Fatty Acid Components of the defensive substances in acid-secreting carabid beetles. Applied Entomology and Zoology. 17(4): 460-466.
59. Kanehisa K, Kawazu K. (1985). Differences in neutral components of the defensive secretion in formic acid-secreting carabid beetles. Applied Entomology and Zoology. 20(3): 299-304.
60. Kanehisa K, Murase M. (1977). Comparative study of the pygidial defensive systems of carabid beetles. Applied Entomology and Zoology 12: 225-235.
61. Katoh K, Standley DM. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution. 30(4): 772-780.
62. Kochevenko A, Araújo WL, Maloney GS, Tieman DM, Do PT, Taylor MG, Klee HJ, Fernie AR. (2012). Catabolism of Branched Chain Amino Acids Supports Respiration but Not Volatile Synthesis in Tomato Fruits. Molecular Plant. 5(2): 366-375.
63. Lange PR, Eastmond PJ, Madagan K, Graham IA. (2004). An *Arabidopsis* mutant disrupted in valine catabolism is also compromised in peroxisomal fatty acid β-oxidation. FEBS Letters. 571(1-3): 147-153.
64. Law CW, Chen Y, Shi W, Smyth GK. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology. 15: R29.
65. Laxme RRS, Suranse V, Sunagar K. (2019). Arthropod venoms: Biochemistry, ecology and evolution. Toxicon. 158: 84-103.
66. Lečić S, Ćurčić S, Vujisić L, Ćurčić B, Ćurčić N, Nikolić Z, Anđelković B, Milosavljević S, Tešević V, Makarov S. (2014). Defensive secretions in three ground-

beetle species (Insecta: Coleoptera: Carabidae). Annales Zoologici Fennici. 51: 285-300.

67. Long JZ, Cravatt BF. (2011). The metabolic serine hydrolases and their functions in mammalian physiology and disease. Chemical Reviews. 111(10): 6022–6063.
68. MacFarlane AJ, Perry CA, Girnary 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. Journal of Biological Chemistry. 284(3): 1533-1539.
69. McCord JM, Fridovich I. (1969). Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). 244(22): 6049-6055.
70. McKenna DD, Shin S, Ahrens D, Balke M, Beza-Beza C, Clarke DJ, Donath A, Escalona HE, Friedrich F, Letsch H, Liu S, Maddison D, Mayer C, Misof B, Murin PJ, Niehuis O, Peters RS, Podsiadlowski L, Pohl H, Scully ED, Yan EV, Zhou X, Ślipiński A, Beutel RG. (2019). The evolution and genomic basis of beetle diversity. Proceedings of the National Academy of Sciences. 116: 24729–24737.
71. McManus R, Ravenscraft A, Moore W (2018) Bacterial associates of a gregarious riparian beetle with explosive defensive chemistry. Frontiers in Microbiology. 9: 2361.
72. Mehler AH, Knox WE. (1950). The conversion of tryptophan to kynurenine in liver. II. The enzymatic hydrolysis of formylkynurenine. Journal of Biological Chemistry. 187(1): 431-438.

73. Meinwald J, Koch KF, Rogers Jr JE, Eisner T. (1966). Biosynthesis of arthropod secretions. III. Synthesis of simple p-benzoquinones in a beetle (*Eleodes longicollis*). Journal of the American Chemical Society. 88: 1590–1592.

74. Meiser J, Schuster A, Pietzke M, Vande Voorde J, Athineos D, Oizel K, Burgos-Barragan G, Wit N, Dhayade S, Morton JP, Dornier E, Sumpton D, Mackay GM, Blyth K, Patel KJ, Niclou SP, Vazquez A. (2018). Increased formate overflow is a hallmark of oxidative cancer. Nature Communications. 9(1): 1368.

75. Meiser J, Tumanov S, Maddocks O, Labuschagne CF, Athineos D, Van Den Broek N, Mackay GM, Gottlieb E, Blyth K, Vousden K, Kamphorst JJ, Vazquez A. (2016). Serine one-carbon catabolism with formate overflow. Scientific Advances. 2(10): e1601273.

76. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution. 37(5): 1530-1534.

77. Moore BP, Brown WV. (1979). Chemical composition of the defensive secretion in *Dyschirius* Bonelli (Coleoptera: Carabidae: Scaritinae) and its taxonomic significance. Journal of the Australian Entomological Society. 18: 123–125.

78. Moore BP, Wallbank BE. (1968). Chemical composition of the defensive secretion in carabid beetles and its importance as a taxonomic character. Proceedings of the Royal Entomological Society of London. 37(5-6): 62-72.

79. Moore BP. (1979). Chemical Defense in Carabids and Its Bearing on Phylogeny. Carabid Beetles, 193–203.

80. Morimoto Y, Bunno S, Fujieda N, Sugimoto H, Itoh S. (2015). Direct Hydroxylation of Benzene to Phenol Using Hydrogen Peroxide Catalyzed by Nickel Complexes Supported by Pyridylalkylamine Ligands. Journal of the American Chemical Society. 137(18): 5867-5870.
81. Moschen I, Bröer A, GalićS, Lang F, Bröer S. (2012). Significance of Short Chain Fatty Acid Transport by Members of the Monocarboxylate Transporter Family (MCT). Neurochemical Research. 37: 2562-2558.
82. Moura A, Savageau MA, Alves R. (2013). Relative amino acid composition signatures of organisms and environments. PloS ONE. 8(10): e77319.
83. Muzzi M, Moore W, Di Giulio A. (2019). Morpho-functional analysis of the explosive defensive system of basal bombardier beetles (Carabidae: Paussinae: Metriini). Micron. 119: 24-38.
84. Oba Y, Yoshida N, Kanie S, Ojika M, Inouye S. (2013). Biosynthesis of firefly luciferin in adult lantern: decarboxylation of L-cysteine is a key step for benzothiazole ring formation in firefly luciferin synthesis. PlOS ONE. 8(12)L e84023.
85. Ober KA, Heider TN. (2010). Phylogenetic diversification patterns and divergence times in ground beetles (Coleoptera: Carabidae: Harpalinae). BMC Evolutionary Biology. 10: 262.
86. Ober KA, Maddison DR. (2008). Phylogenetic relationships of tribes within Harpalinae (Coleoptera: Carabidae) as inferred from 28S ribosomal DNA and the wingless gene. Journal of Insect Science. 8: 63.

87. Ortiz de Montellano PR. (2010). Hydrocarbon hydroxylation by cytochrome P450 enzymes. Chemical Reviews. 110(2): 932–948.

88. Pamin K, Połtowicz J, Prończuk M, Basąg S, Maciejewska J, Kryściak-Czerwenka J, Tokarz-Sobieraj R. (2015). Hydroxylation of phenol by hydrogen peroxide catalyzed by heteropoly compounds in presence of glycerol as green solvent. Catalysis Today. 257(1): 80-85.

89. Paradis E, Claude J, Strimmer K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 20(2): 289-290.

90. Petersen M, Meusemann K, Donath A, Dowling D, Liu S, Peters RS, Podsiadlowski L, Vasilikopoulos A, Zhou X, Misof B, Niehuis O. (2017). Orthograph: a versatile tool for mapping coding nucleotide sequences to clusters of orthologous genes. BMC Bioinformatics. 18: 111.

91. Punta M, Coggill 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 Research. 40: D290-301.

92. 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. Proceedings of the National Academy of Sciences of the United States of America. 109(37): 14858-1463.

93. Quistgaard EM, Löw C, Guettou F, Nordlund P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. Nature Reviews Molecular Cell Biology. 17: 123-132.

94. Raspotnig G, Schaider M, Föttinger P, Leutgeb V, Komposch C. (2015). Benzoquinones from scent glands of phalangiid harvestmen (Arachnida, Opiliones, Eupnoi): a lesson from Rilaena triangularis. Chemoecology. 25: 63–72.

95. 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 Research. 43(7): e47.

96. Rocha DFO, Wouters FC, Zampieri DS, Brocksom TJ, Machado G, Marsaioli AJ. (2013). Harvestman phenols and benzoquinones: characterisation and biosynthetic pathway. Molecules 18: 11429–11451.

97. Rodriguez J, Jone TH, Sierwal P, Marek PE, Shear WA, Brewer MS, Kocot KM, Bond JE. (2018). Step-wise evolution of complex chemical defenses in millipedes: a phylogenomic approach. Scientific Reports. 8: 3209.

98. Rork AM, Mikó I, Renner T. (2019). Pygidial glands of *Harpalus pensylvanicus* (Coleoptera: Carabidae) contain resilin-rich structures. Arthropod Structure & Development. 44(12): 1069-1083.

99. Rork AM, Renner T. (2018). Carabidae Semiochemistry: Current and Future Directions. Journal of Chemical Ecology. 49: 19-25.

100. Rork AM, Xu S, Attygalle A, Renner T. (2020). Primary Metabolism Co-opted for Defensive Chemical Production in the Carabid Beetle, *Harpalus pensylvanicus*. Journal of Chemical Ecology. 47(3): 334-349.

101. Rossini C, Attygalle AB, González A, Smedley SR, Eisner M, Meinwald J, Eisner T. (1997). Defensive production of formic acid (80%) by a carabid beetle (*Galerita*

lecontei). Proceedings of the National Academy of Sciences of the United States of America. 94: 6792-6797.

102. Roth LM, Eisner T. (1962). Chemical Defenses of Arthropods. Annual Review of Entomology. 7: 107-136.

103. 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. Biochemical Pharmacology. 34(3): 361-367.

104. Sayols S (2020). rrvgo: a Bioconductor package to reduce and visualize Gene Ontology terms.

105. Schildknecht H, Maschwitz U, Winkler H. (1968). Zur Evolution der Carabiden-Wehrdrüsensekrete. Über Arthropoden-Abwehrstoffe XXXII. Naturwissenschaften. 55: 112–117.

106. Schildknecht H. (1964). Defensive substances of the Arthropods, their isolation and identification. Angewandte Chemie International Edition. 3(2): 73-82

107. Schildknecht H. (1970). The Defensive Chemistry of Land and Water Beetles. Angewandte Chemie International Edition. 9(1): 1-9.

108. Schlosser PM, Bond JA, Medinsky MA. (1993). Benzene and phenol metabolism by mouse and rat liver microsomes. Carcinogenesis. 14(12): 2477–2486.

109. Schuler MA, Berenbaum MR. (2013). Structure and function of cytochrome P450s in insect adaptation to natural and synthetic toxins: insights gained from molecular modeling. Journal of Chemical Ecology. 39(9): 1232-1245.

110. Scott JG, Wen Z. (2001). Cytochromes P450 of insects: the tip of the iceberg. Pest Management Science. 57(10): 958-967.

111. Sekowska A, Ashida H, Danchin A. (2019). Revisiting the methionine salvage pathway and its paralogues. Microbial Biotechnology. 12(1): 77-97.

112. 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.

113. Storey JD. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B. 64: 479–498.

114. Sützl L, Foley G, Gillam E, Bodén M, Haltrich D. (2019). The GMC superfamily of oxidoreductases revisited: analysis and evolution of fungal GMC oxidoreductases. Biotechnology for Biofuels. 12: 118.

115. Sygmund C, Klausberger M, Felice AK, Ludwig R. (2011). Reduction of quinones and phenoxy radicals by extracellular glucose dehydrogenase from *Glomerella cingulata* suggests a role in plant pathogenicity. Microbiology. 157(11): 3203–3212.

116. Symonds MRE and Elgar MA. (2003). The mode of pheromone evolution: evidence from bark beetles. Proceedings of the Royal Society B: Biological Sciences. 271(1541): 839-846.

117. The Gene Ontology Consortium. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics. 25: 25-29.

118. van Berkel WJ, Eppink MH, Middelhoven WJ, Vervoort J, Rietjens IM. (1994). Catabolism of 4-hydroxybenzoate in *Candida parapsilosis* proceeds through initial

oxidative decarboxylation by a FAD-dependent 4-hydroxybenzoate 1-hydroxylase. FEMS microbiology Letters. 121(2): 207–215.

119. Vasilikopoulos A, Balke M, Kukowka S, Pflug JM, Martin S, Meusemann K, Hendrich L, Mayer C, Maddison DR, Niehuis O, Beutel RG, Misof B. (2021). Phylogenomic analyses clarify the pattern of evolution of Adephaga (Coleoptera) and highlight phylogenetic artefacts due to model misspecification and excessive data trimming. Systematic Entomology. 46(4): 991-1018.

120. Wanders RJA, Duran M, Loupatty FJ. (2010). Enzymology of the branched-chain amino acid oxidation disorders: the valine pathway. Journal of Inherited Metabolic Disease. 35: 5-12.

121. Wang W, An C, Yao Y, Meng X, Gao SS. (2021). De novo biosynthesis and gram-level production of m-cresol in *Aspergillus nidulans*. Applied Microbiology and Biotechnology. 105(16-17): 6333–6343.

122. Will KW, Attygalle AB, Herath K. (2000). New defensive chemical data for ground beetles (Coleoptera: Carabidae): interpretations in a phylogenetic framework. Biological Journal of the Linnean Society. 71: 459-481.

123. Wu Q, Chen X, Li J, Sun S. (2020). Serine and Metabolism Regulation: A Novel Mechanism in Antitumor Immunity and Senescence. Aging and Disease. 11(6): 1640–1653.

124. Yan Z, Zhong HM, Maher N, Torres R, Leo GC, Caldwell GW, Huebert N. (2005). Bioactivation of 4-methylphenol (p-cresol) via cytochrome P450-mediated aromatic oxidation in human liver microsomes. Drug Metabolism and Disposition: The Biological Fate of Chemicals. 33(12): 1867–1876.

125. Young MD, Wakefield MJ, Smyth GK, Oshlack A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 11: R14.

126. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 8(1): 28-36.

127. Zheng Y, Ramsamooj S, Li Q, Johnson JL, Yaron TM, Sharra K, Cantley LC. (2019). Regulation of folate and methionine metabolism by multisite phosphorylation of human methylenetetrahydrofolate reductase. Scientific Reports. 9: 4190.

FIGURES AND TABLES

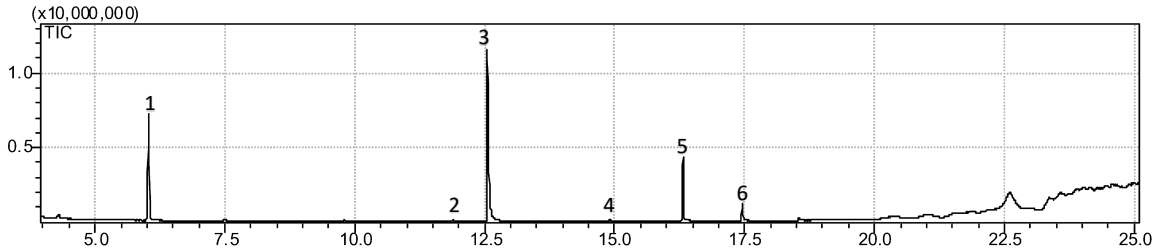


Figure 1. Gas chromatogram of the pygidial gland sprays of *Platynus angustatus*. In order of occurrence, the labelled peaks correspond to the following compounds: 1) Undecane 2) acetic acid 3) formic acid 4) 2-tridecanone 5) 2-tetradecanone 6) 2-pentadecanone.

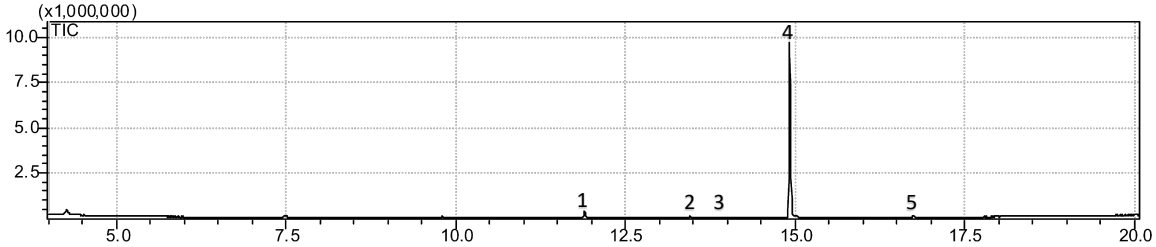


Figure 2. Gas chromatogram of the pygidial gland sprays of *Pterostichus moestus*. In order of occurrence, the labelled peaks correspond to the following compounds: 1) Acetic acid 2) propanoic acid 3) 2-methylpropanoic acid 4) methacrylic acid 5) tiglic acid.

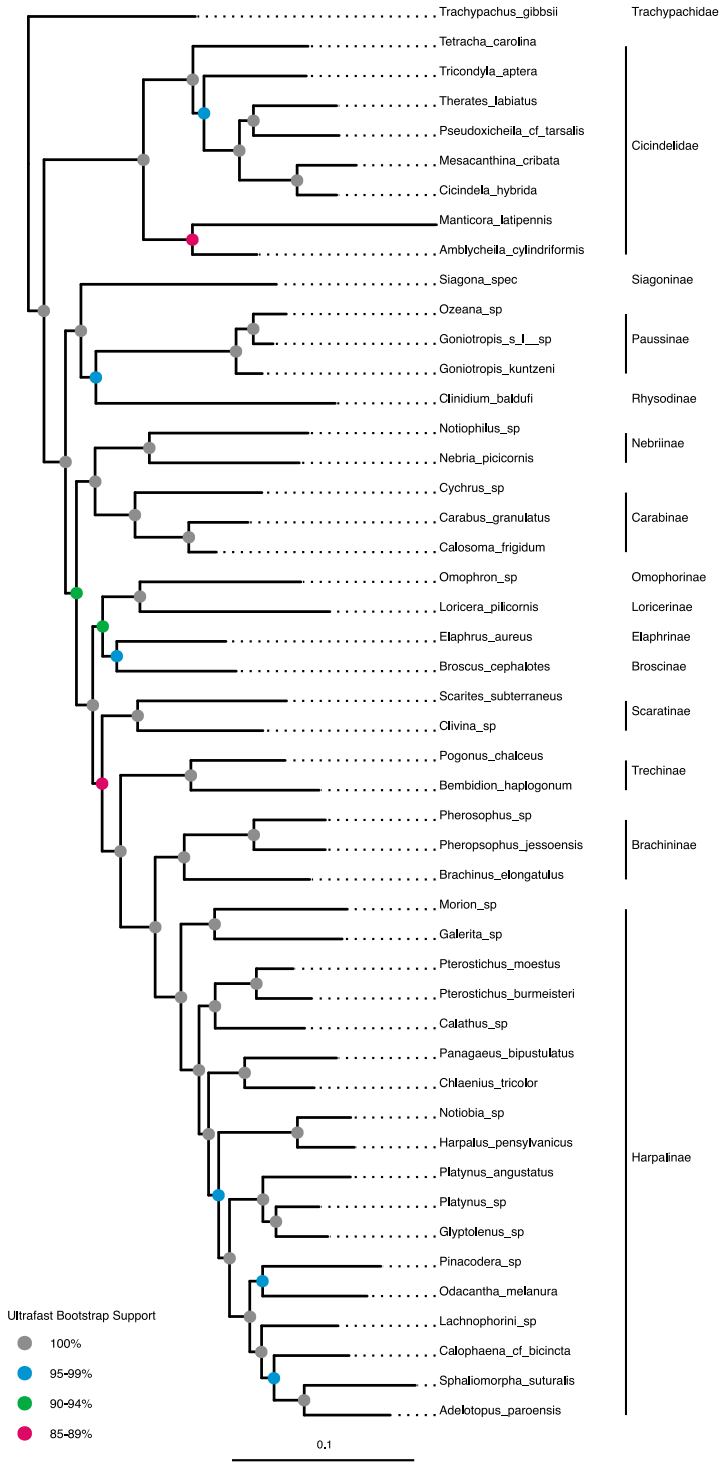


Figure 3. Phylogenetic reconstruction of Geadephaga (Carabidae + Cicindelidae + Trachypachidae). Subfamily labels are shown to the right of the phylogeny. Ultrafast

bootstrap support values (2000 replicates) have been condensed to a four-color scheme: gray nodes = 100% BS, blue nodes = 95-99% BS, green nodes = 90-94% BS, vermillion nodes = 85-89% BS.

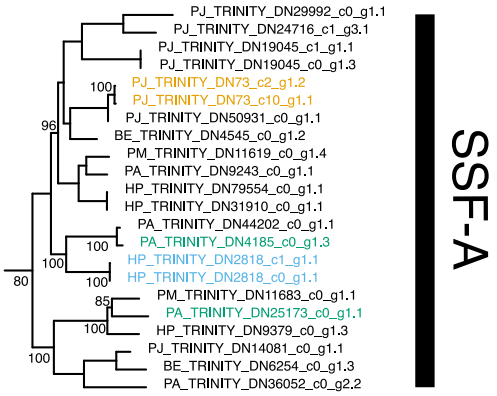


Figure 4a. Clade SSF-A, subsetted from the larger solute symporter family phylogeny. Genes are labelled with their species initials and genes upregulated in the secretory lobes of have been colored in a species-specific manner. Ultrafast bootstrap support values have been removed from all nodes except for those most important to understanding the relatedness of upregulated sequences. Species initials are as follows: BE=*Brachinus elongatulus*; GK=*Goniotropis kuntzeni*; HP=*Harpalus pensylvanicus*; PA=*Platynus angustatus*; PJ=*Pheropsophus jessoensis*; PM=*Pterostichus moestus*.

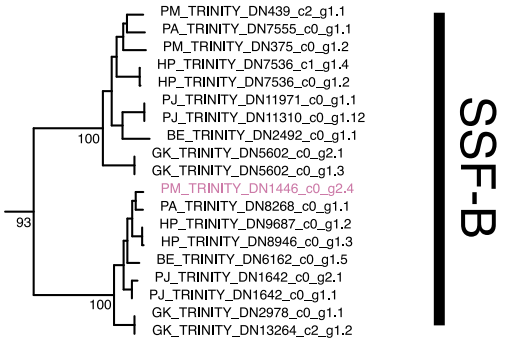


Figure 4b. Clade SSF-B, subsetted from the larger solute symporter family phylogeny. Genes are labelled with their species initials and genes upregulated in the secretory lobes of have been colored in a species-specific manner. Ultrafast bootstrap support values have been removed from all nodes except for those most important to understanding the relatedness of upregulated sequences. Species initials are as follows: BE=*Brachinus elongatulus*; GK=*Goniotropis kuntzeni*; HP=*Harpalus pensylvanicus*; PA=*Platynus angustatus*; PJ=*Pheropsophus jessoensis*; PM=*Pterostichus moestus*.

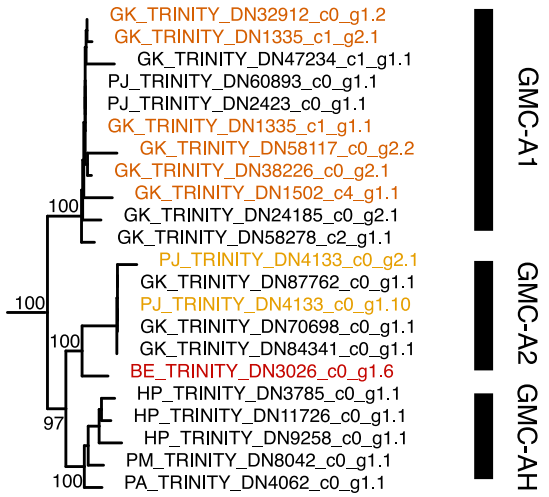


Figure 5a. Clade GMC-A, subsetted from the larger GMC oxidoreductase family phylogeny. Genes are labelled with their species initials and genes upregulated in the secretory lobes of have been colored in a species-specific manner. Ultrafast bootstrap support values have been removed from all nodes except for those most important to understanding the relatedness of upregulated sequences. Species initials are as follows: BE=*Brachinus elongatulus*; GK=*Goniotropis kuntzeni*; HP=*Harpalus pensylvanicus*; PA=*Platynus angustatus*; PJ=*Pheropsophus jessoensis*; PM=*Pterostichus moestus*.

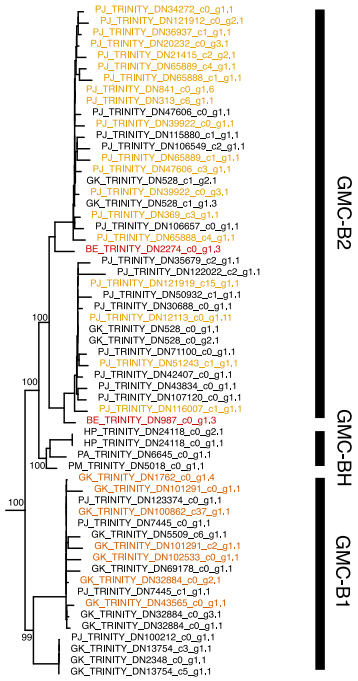


Figure 5b. Clade GMC-B, subsetted from the larger GMC oxidoreductase family phylogeny. Genes are labelled with their species initials and genes upregulated in the secretory lobes of have been colored in a species-specific manner. Ultrafast bootstrap support values have been removed from all nodes except for those most important to understanding the relatedness of upregulated sequences. Species initials are as follows: BE=*Brachinus elongatulus*; GK=*Goniotropis kuntzeni*; HP=*Harpalus pensylvanicus*; PA=*Platynus angustatus*; PJ=*Pheropsophus jessoensis*; PM=*Pterostichus moestus*.

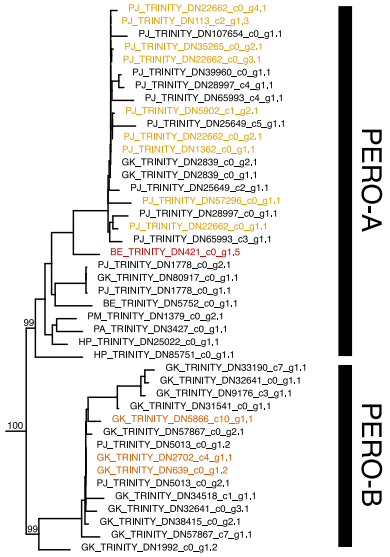


Figure 6. Clade PERO-A/B, subsetted from the larger animal haem peroxidase phylogeny. Genes are labelled with their species initials and genes upregulated in the secretory lobes of have been colored in a species-specific manner. Ultrafast bootstrap support values have been removed from all nodes except for those most important to understanding the relatedness of upregulated sequences. Species initials are as follows: BE=*Brachinus elongatulus*; GK=*Goniotropis kuntzeni*; HP=*Harpalus pensylvanicus*; PA=*Platynus angustatus*; PJ=*Pheropsophus jessoensis*; PM=*Pterostichus moestus*.

	Brachinus elongatulus	*Goniotropis kuntzeni*	*Harpalus pensylvanicus*	*Platynus angustatus*	*Pheropsophus jessoenis*	*Pterostichus moestus*
Reads (pre-trim)	1.04E+08	4.23E+08	3.00E+08	5.49E+08	4.46E+08	4.24E+08
Reads Out (post-trim)	1.01E+08	4.15E+08	2.93E+08	5.38E+08	4.38E+08	4.09E+08
Percent Reads Retained	96.8%	98.2%	97.8%	98.1%	98.3%	96.4%
Bases In (pre-trim)	1.04E+10	6.34E+10	4.53E+10	8.26E+10	6.68E+10	6.39E+10
Bases Out (post-trim)	1.01E+10	6.21E+10	4.27E+10	7.17E+10	6.52E+10	5.43E+10
Percent Bases Retained	96.5%	97.9%	94.3%	86.8%	97.6%	84.9%

Table 1. Summary statistics on raw sequencing data output and data retention post-trimming for each species.

	Brachinus elongatulus	*Goniotropis kuntzeni*	*Harpalus pensylvanicus*	*Platynus angustatus*	*Pheropsophus jessoenis*	*Pterostichus moestus*
Genes	21,688	176,428	93,716	63,405	141,250	49,008
Transcripts	49,340	253,753	163,238	145,296	219,405	105,544
Transcript N50	2,581 bp	1,573 bp	1,621 bp	1,613 bp	1,625 bp	1,837 bp
Transcript E90N50	2,096 bp	2,036 bp	1,750 bp	1,649 bp	1,738 bp	1,682 bp
GC Content	36.4%	36.4%	36.9%	36.3%	38.4%	37.3%
Assembly Size	8.30E+07 bp	2.52E+08 bp	1.67E+08 bp	1.57E+08 bp	2.26E+08 bp	1.23E+08 bp

Table 2. Summary statistics on *de novo* transcriptome assembly size, transcript contiguity, and gene content for each species.

	Brachinus elongatulus	*Goniotropis kuntzeni*	*Harpalus pensylvanicus*	*Platynus angustatus*	*Pheropsophus jessoenis*	*Pterostichus moestus*
BUSCO Complete	98.3%	99.5%	97.4%	98.5%	99.5%	98%
BUSCO Single	39.3%	7.7%	4.6%	12.3%	8.6%	33.9%
BUSCO Duplicated	59%	91.8%	92.8%	86.2%	90.9%	64.1%
BUSCO Fragmented	0.9%	0.2%	0.9%	0.6%	0.1%	0.9%
BUSCO Missing	0.8%	0.3%	1.7%	0.9%	0.4%	1.1%

Table 3. Summary statistics on BUSCO completeness for each species. BUSCO Single refers to the percent of benchmark orthologs in the transcriptome which are present only once. BUSCO Duplicated refers to the percent of benchmark orthologs in the transcriptome which are present twice or more. Inflated "Duplicated" scores as seen here

are typical for transcriptomes given that BUSCO will consider multiple isoforms of a sequence as duplication.

	Brachinus elongatulus	*Goniotropis kuntzeni*	*Harpalus pensylvanicus*	*Platynus angustatus*	*Pheropsophus jessoenis*	*Pterostichus moestus*
Transcripts w/ CDS	33,525	113,362	75,652	65,054	122,792	50,311
Coding Sequences	40,619	131,530	86,399	69,569	144,605	55,837
Transcripts w/ BlastX Hit	26,523	98,867	57,444	47,638	97,804	37,740
Transcripts w/ BlastP Hit	29,567	79,466	57,106	44,472	90,971	36,654
Transcripts w/ HMMScan Hit	26,983	70,388	52,161	41,949	83,629	33,362

Table 4. Summary statistics on coding sequence prediction and functional annotation for each species. BLAST searches were run against the UniProtKB database whereas HMMScan searches were run against the Pfam-A database.

	Brachinus elongatulus	*Goniotropis kuntzeni*	*Harpalus pensylvanicus*	*Platynus angustatus*	*Pheropsophus jessoenis*	*Pterostichus moestus*
Avg. percent pseudoalignment	97.2%	89.3%	68/0%	75.8%	85.8%	81.9%
DE genes	2,488	2,497	1,479	1,198	4,705	4,358
Genes SLUP	286	451	1,010	469	522	488
Genes RCUP	163	245	NA	NA	362	NA

Table 5. Summary statistics on the pseudoalignment rates of reads back to their respective transcriptome assemblies and differential gene expression analyses for each species.

CHAPTER 5

Dissertation Synthesis

ABSTRACT

Exocrine glands are ubiquitous across the Arthropoda and highly variable in several respects, including but not limited to their morphologies, physiologies, biochemistries, evolutionary histories, and ecological significance. Particularly widespread are exocrine glands used for self-defensive purposes which have been found in every arthropod body region from the tips of antennae to the terminal-most abdominal segment. This work primarily focuses on the pygidial defensive glands of Carabidae with particular emphasis on elucidating the molecular mechanisms underlying defensive chemical synthesis and identifying morphological adaptations to chemical defense storage and secretion. Although this dissertation was narrow in taxonomic scope, I suggest that the general approaches taken to examining these topics can be extended to studying the form, function, and evolution of exocrine glands across the Arthropoda. Given the pivotal role that these tissues play in the attraction of mates, repulsion of predators, foraging behavior, etc., I strongly encourage their study such that we can better understand not only the evolution of these structures for their own sake, but to further our understanding of the evolution these important behavioral phenotypes and the taxa which possess them.

SYNTHESIS

Exocrine gland structure and function are codependent phenotypes. Indeed, the capacity of gland cells to produce semiochemicals has little ecological significance without having the infrastructure to transport, store, and secrete them. The capacity of gland systems to transport, store, and secrete such semiochemicals likewise means little without compounds to release into the environment. Thus, to fully understand how gland systems evolve and function, we must not approach their biochemistries, their transcriptional profiles, ecological roles, or their morphological features as though they are closed systems; we must approach and understand them as intricately connected aspects of a larger whole. I propose that the approaches taken throughout this dissertation, despite having been used entirely within the family Carabidae, are broadly applicable and are well-suited to working toward such an aim for nearly all arthropod exocrine gland systems.

My second chapter was primarily focused on examining gland structure, namely the morphological adaptations to allomone storage in the pygidial glands of the carabid beetle, *Harpalus pensylvanicus* (Rork et al. 2019). This work was conducted using a combination of confocal laser scanning microscopy and histological techniques. Indeed, confocal laser scanning microscopy's primary utility in arthropod morphology depends upon the autofluorescent properties of arthropod tissue, properties which have been characterized and shown to be consistent across all major lineages of Arthropoda (Michels et al. 2016). Thus, the methods used in this chapter could theoretically be used to study silk glands in Araneae, cyanide glands in Xystodesmidae, Dufour's glands in Apocrita, etc. The same is generally true of the histological methods used, which could

be used to complement or expand upon the ability of CLSM to detect differences in tissue composition via autofluorescence (Michels et al. 2016). Of course, I would also highly encourage the use of other methods for studying gland structure such as scanning electron microscopy, particularly that paired with ablation instrumentation which would allow for better understanding of gland ultrastructure (Di Giulio et al. 2015).

Also broadly applicable to the study of others lineages' gland systems is the comparative transcriptomics approach I took in chapters three and four to identify putative pathways involved in semiochemical biosynthesis. The utility of this method relies upon the reasonable assumption that genes significantly upregulated in gland tissue relative to all other tissues may have some important role in gland function. Given this, this approach has tremendous power to detect not only genes involved in semiochemical biosynthesis roles, but other roles such as gland development, semiochemical transport, the maintenance of gland homeostasis, etc. I urge that this is not only true for taxa with relatively large glands from which relatively large quantities of RNA can be extracted. With the advent of technologies such as single-cell RNA-Seq and cDNA library preparation methods allowing for picogram inputs of RNA, even the smallest arthropods with the smallest glands can be studied with this method assuming gland excision itself is a practical task (Schuierer et al. 2017, Sheng et al. 2020).

This all leads to a practical and important question: Why study the structure, function, and evolution of exocrine glands? I stress that an understanding of gland evolution is essential to fully understanding the evolution of lineages possessing such glands. This is necessarily true, as exocrine glands would definitionally be components of the phenome, and to work toward a full understanding of the evolution of any organism

means working toward understanding the evolution of all its phenotypes. Of course, it is also evident that exocrine glands do not function or evolve in a closed system as previously alluded to. As discussed in my first chapter, exocrine glands produce some of the most important biochemicals in arthropods, from sex pheromones required for mating success to defensive chemicals important for self-defense, from brood pheromones underpinning colony dynamics in eusocial bees to kairomones necessary for mediating parasitoid-host interactions (Rork & Renner 2018). They are in fact the source of most chemically mediated behaviors, behaviors which have established roles in the evolution of lineages as phenotypes upon which natural selection acts (Ebert & Fields 2020, Jones & Ratterman 2009). While I do not suggest that an understanding of behavioral evolution necessitates an understanding of gland evolution or vice-versa, I do stress their interrelatedness. Chemically mediated behaviors would not exist without chemicals to mediate them, thus making exocrine glands necessary for the evolution of most if not all chemically mediated behaviors. Glandular phenotypes are in turn acted upon by natural selection, which may favor the biosynthesis of stronger defensive chemical blends, different enantiomeric ratios in a pheromone blend, or reduced kairomone output as examples (Weiss et al. 2017). Thus, exocrine glands and chemically mediated behaviors are complementary areas of inquiry and likely coevolve, making the understanding of each part necessary for understanding the larger whole. Indeed, exocrine gland evolution is not only fascinating for its own sake, but an understanding of it is requisite to fully understanding the evolution of chemically mediated behavior and vice-versa, both of which are vital to fully understanding the evolution of lineages broadly.

This dissertation can be seen as validation of these points. I have demonstrated that **the process of co-option likely plays a critical role in the evolution of chemical defense in the Carabidae**. Specifically, formic acid and methacrylic acid are likely biosynthesized in the secretory lobes via the co-option of pathways and gene regulatory networks that long predate this family's crown age (c. 200 mya) (Baca et al. 2021, Rork et al. 2021). We have also provided evidence that while bombardier beetles may not biosynthesize benzoquinones via a canonical pathway, they have likely co-opted members of ancient gene families for the process. This underscores the important point that co-option may not only be a prominent mechanism of biochemical evolution, but a key driver in many lineages (Gould & Vrba 1982). I have also demonstrated that **the process of co-option likely plays a role in the evolution of morphological adaptations to defensive chemical storage in the Carabidae**. Given its primary role as a connective tissue across the Arthropoda, the discovery of resilin as a (chemically resistant) duct liner in the pygidial glands can be treated as an exaptation (Gould & Vrba 1982, Rork et al. 2019). That is, this protein initially evolved as a connective tissue given its mechanically resistant and elastomeric properties, but in Carabidae was co-opted as a duct liner in the pygidial glands. Given its chemically resistant properties, this likely would be favored by natural selection (Chen & Thouas 2014). I suggest that the co-option of mechanically and chemically resistant materials to form the glands themselves is likely widespread, especially in those glands which store relatively reactive compounds (Rork et al. 2019). **Taken together, I suggest that the primary defense mechanism of the Carabidae appears to evolve in large part through the process of co-option rather than *de novo* mechanisms.**

What this implies for broader Carabidae biology and evolution will certainly take further investigation and consideration, but we can derive a few key points from these conclusions. Amongst the most significant in my view is an explanation as to why we see relatively few primary defensive compounds evolving multiple times across the family, and why we see practically no novel defensive compounds (Rork & Renner 2018). It seems at first remarkable that of the hundreds of unique compounds characterized from hundreds of species across the family, none are entirely novel. This is especially interesting in the broader context of coleopteran chemical defense, where the abundance of novel defensive chemicals in other lineages such as Melodidae (cantharidin), Staphylinidae (stenusin), and even Gyrinidae (gyrinidal) is relatively common (Dettner 1987). However, if co-option truly is the primary driving mechanism of defensive chemical evolution, this lack of novelty is entirely to be expected. That is, most plausible evolutionary trajectories through which defensive chemicals would evolve would be via the co-option of some evolutionarily conserved pathway producing a similarly conserved metabolite. This of course raises more questions such as why co-option would be the primary evolutionary mechanism underlying chemical defense in the first place, why we don't see more novel enzyme functions and pathways evolving that produce repellent compounds, etc., but these questions begin to fall outside the scope of this synthesis. We unfortunately do not know nearly enough about chemical defense in the context of natural history in the Carabidae to draw many conclusions about the implications of this work to that field, a point articulated at length in my first chapter (Rork & Renner 2018). In theory, though, having answers to relatively simple questions like "For some carabid species' primary predator(s), are certain compounds more efficacious repellents than

others?" could help us understand how selection by predators shapes the evolution of chemical defense, if at all, and how defensive chemistry shapes predator-prey relationships across the Carabidae. It cannot be understated how fundamental such a question is, as predator-prey dynamics have been shown to play a substantial role in the evolution of lineages across life (Hauge et al. 2020, Heiling & Herberstein 2004, Nair et al. 2019, Wilson et al. 2018).

How exactly such inquiries into exocrine gland evolution would apply to other systems is of course highly dependent on the gland system, the role it and its semiochemicals play in mediating behavior, and several other factors. Perhaps we could unveil mechanisms underlying the switch from venom to formic acid production in formicine ants or molecular drivers of pheromone diversity and divergence across Lepidoptera. Perhaps by learning enough about the mechanisms underlying the evolution of glands, we can begin to understand why they are seemingly abundant in some lineages and relatively lacking in others. In any case, exocrine glands and their evolution should be seen, in my view, as a rich resource by which we can begin to better understand not only evolutionary mechanisms for their own sake, but how these mechanisms shape the evolution of life as a whole.

CONCLUSIONS

Exocrine glands represent a fascinating array of morphologies, developmental origins, physiologies, biochemistries, ecological functions, and evolutionary trajectories. Indeed, they are some of life's most important biochemical factories, necessary for producing the chemical cues and signals that underlie its most astounding and important behavioral

phenotypes. Given this vital role, an understanding of glandular form, function, and evolution should be viewed as a prerequisite to truly understanding the evolution of the lineages possessing them. In this dissertation, we have demonstrated through the study of exocrine gland evolution in the Carabidae how one of the animal kingdom's most diverse families is capable of synthesizing, storing, and secreting powerful repellents, from astringent acids to boiling benzoquinones. Through this work and work to come, we can begin contextualizing the evolution of this family's most unique phenotypes to better understand how the Carabidae as a lineage have evolved. We urge the entomological and evolutionary biology communities that our general approach can be extended to nearly any system and that studies into exocrine gland evolution can be undertaken not only across the phylum Arthropoda, but across the Metazoa. Only through this effort can we truly unveil the mechanisms driving the evolution of chemically mediated behaviors, from mating to defense to parasitism, thus allowing us to better understand the evolution of these taxa more broadly.

ACKNOWLEDGEMENTS

I would like to thank Tanya Renner, Julie Urban, Etya Amsalem, and Shaun Mahony for their input and helpful discussion regarding the structure of this synthesis. This work was supported by the National Science Foundation (DEB 1762760) and The Coleopterists Society Graduate Student Research Enhancement Award. The findings and conclusions of this work do not necessarily reflect the views of the National Science Foundation or The Coleopterists Society.

REFERENCES

1. Baca SM, Gustafson GT, Alexander AM, Gough HM, Toussaint EFA. (2021). Integrative phylogenomics reveals a Permian origin of Adephaga beetles. Systematic Entomology. 46(4): 968-990.
2. Chen Q, Thouas G. (2014). Chapter 10: Bioresorbable Polymers. In: Biomaterials: A Basic Introduction. CRC Press, Taylor Francis Group, Florida, pp. 360-361.
3. Dettner K. (1987). Chemosystematics and evolution of beetle chemical defenses. Annual Review of Entomology. 32: 17-48.
4. Di Giulio A, Muzzi M, and Romani R. (2015). Functional anatomy of the explosive defensive system of bombardier beetles (Coleoptera, Carabidae, Brachininae). Arthropod Structure and Development. 44(5): 468-490.
5. Ebert D, Fields PD. (2020). Host-parasite co-evolution and its genomic signature. Nature Reviews Genetics. 21: 754-768.
6. Gould SJ, Vrba ES. (1982). Exaptation—a missing term in the science of form. Paleobiology. 8(1): 4-15.
7. Hague MTJ, Stokes AN, Feldman CR, Brodie ED Jr, Brodie ED III. (2020). The geographic mosaic of arms race coevolution is closely matched to prey population structure. Evolution Letters. 4(4): 317-332.
8. Heiling AM, Herberstein ME. (2004). Predator-prey coevolution: Australian native bees avoid their spider predators. Proceedings of the Royal Society B. 271 Suppl. 4(Suppl. 4): S196–S198.

9. Jones AG, Ratterman NL. (2009). Mate choice and sexual selection: What have we learned since Darwin? Proceedings of the National Academy of Sciences of the United States of America. 106 Suppl. 1(Suppl. 1): 10001–10008.
10. Michels, J., Appel, E. and Gorb, S.N., 2016. Functional diversity of resilin in Arthropoda. Beilstein Journal of Nanotechnology. 7: 1241-1259.
11. Nair RR, Vasse M, Wielgoss S, Sun L, Yu YN, Velicer GJ. (2019). Bacterial predator-prey coevolution accelerates genome evolution and selects on virulence-associated prey defences. Nature Communications. 10: 4301.
12. Rork AM, Mikó I, Renner T. (2019). Pygidial glands of *Harpalus pensylvanicus* (Coleoptera: Carabidae) contain resilin-rich structures. Arthropod Structure & Development. 44(12): 1069-1083.
13. Rork AM, Renner T. (2018). Carabidae Semiochemistry: Current and Future Directions. Journal of Chemical Ecology. 49: 19-25.
14. Rork AM, Xu S, Attygalle A, Renner T. (2021). Primary Metabolism co-Opted for Defensive Chemical Production in the Carabid Beetle, *Harpalus pensylvanicus*. Journal of Chemical Ecology. 47(3): 334-349.
15. Schuierer S, Carbone W, Knehr J, Petitjean V, Fernandez A, Sultan M, Roma G. (2017). A comprehensive assessment of RNA-seq protocols for degraded and low-quantity samples. BMC Genomics. 18: 422.
16. Sheng L, Shields EJ, Gospocic J, Glastad KM, Ratchasanmuang P, Berger SL, Raj A, Little S, Bonasio R. (2020). Social reprogramming in ants induces longevity-associated glia remodeling. Science Advances. 6(34): eaba9869.

17. Weiss K, Herzner G, Strohm E. (2017). Sexual selection and the evolution of male pheromone glands in philanthine wasps (Hymenoptera, Crabronidae). BMC Evolutionary Biology. 17: 128.
18. Wilson AM, Hubel TY, Wilshin SD, Lowe JC, Lorenc M, Dewhirst OP, Bartlam-Brooks HLA, Diack R, Bennitt E, Golabek KA, Woledge RC, McNutt JW, Curtin NA, West TG. (2018). Biomechanics of predator–prey arms race in lion, zebra, cheetah and impala. Nature. 554: 183-188.

APPENDIX

Chapter 3 Supplementary Captions

This work has been previously published and has been reproduced with permission from Springer Nature (https://doi.org/10.1007/s10886-021-01253-2).

Supplementary Figure 3-1a-c
Mass spectra of formic acid (A), 2-tridecanone (B), and undecane (C) obtained from GC-MS analysis of the pygidial gland sprays of *Harpalus pensylvanicus*. The formic acid spectrum was obtained via the Shimadzu QP5050 MS while the 2-tridecanone and undecane mass spectra were obtained via an Agilent 6890-5973 GC-MS.

Supplementary Figure 3-2
Simplified representation of a portion of the kynurenine pathway. 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. Tryptophan 2,3-dioxygenase (TDO), a non-differentially expressed gene in the secretory lobes of *H. pensylvanicus*, oxidizes L-tryptophan to form N'formylkynurenine. The arylamine bond of N'formylkynurenine is hydrolyzed by kynurenine fomamidase, generating formic acid and L-kynurenine.

Supplementary Figure 3-3
Highly simplified representation of a portion of the methionine salvage cycle. The circle labelled "Multiple intermediate steps" is shown for readability. 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. S-adenosylmethionine is formed from L-methionine and ATP via methionine adenosyltransferase (MAT). From here, a number of intermediate steps occur prior to the formation of 1,2-dihydroxy-3-keto-5-methylthiopentene, as well as multiple shunts from the salvage cycle. 1,2-dihydroxy-3-keto-5-methylthiopentene is oxidized by 1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase (ADI1), generating formic acid and 2-keto-4-methylthiobutyrate. A transamination reaction is then catalyzed between 2-keto-4-methylthiobutyrate and L-glutamate via the enzyme L-glutamine-4-(methylsulfanyl)-2-oxobutanoate aminotransferase (mtnE), regenerating L-methionine and forming 2-oxoglutarate.

Supplementary Figure 3-4a-d
(A) Sample correlation matrix generated using log2-transformed expression data from the differentially expressed genes. Cluster trees and matrix coloration show more intrasample correlation between replicates than intersample correlation. (B) A differentially expressed gene vs. samples heat map generated using log2-transformed expression data from the differentially expressed genes. Each row represents one gene, showing higher correlation intrasample correlation between replicates than intersample correlation for most genes. (C) MA plot showing significantly differentially expressed genes (logFC > 4, adj.p-value < 0.05) in red and non-differentially expressed genes in black. (D) Volcano plot showing

significantly differentially expressed genes (logFC > 4, adj.p-value < 0.05) in red and non-differentially expressed genes in black.

Supplementary Table 3-1
Statistics on raw sequencing data output, data retained post-trimming, and data discarded via trimming. Length cutoff was set at 100bp, quality cutoff at Q20, and complexity cutoff at 30% complexity.

Supplementary Table 3-2
Transcriptome assembly and quality control statistics (reported as BUSCO v3 scores using the endopterygota database).

Supplementary Table 3-3
Transcriptome annotation statistics. BLASTX and BLASTP hits were against the UniProtKB database. HMMScan hits were against the Pfam-A database.

Supplementary Table 3-4
The transcriptome annotation database output by the Trinotate pipeline. Columns 1-13 (A-M in Excel) contain the regular output of Trinotate with the exclusion of the signalP, tmHMM, and RNAMMER columns. Columns 14-21 (N-U in Excel) contain replicate-wise TMM-normalized TPM values for each gene as well as the average of said values per gene per tissue type. Columns 22-25 (V-Y in Excel) contain data pertaining to the differential gene expression analyses carried out between the secretory lobe and whole body samples. Only genes which are significantly upregulated in the secretory lobes have values shown here.

Supplementary Table 3-5
A list of all GO Terms enriched in the secretory lobes of Harpalus pensylvanicus relative to glandless body tissue. Included in this table are the GO categories and their corresponding terms and ontologies. Over- and under-represented p-values are also included along with over-represented p-values adjusted via the Benjamini-Hochberg procedure. The total number of genes corresponding to each term are similarly included, along with the number of upregulated genes in said term and their identities.

Supplementary Table 3-6
A tab-separated index file used for curation of the sequence database used for the amorphean mTHF phylogeny. Column 1 contains all NCBI Accessions, column 2 contains the sequence's gene classification, and column 3 contains the species to which each sequence belongs. Also in this table are the two *Harpalus pensylvanicus* TRINITY IDs incorporated in the phylogeny.

Supplementary Data 3-1
Raw sequence data, amino acid and coding sequence alignments, and results of the maximum-likelihood phylogenetic reconstruction of the amorphean mTHF gene family as well as positive episodic diversifying selection tests.

Supplementary File 3-1
Link to interview with Trinity RNA-Seq developer Brian Hass regarding the preferred read length to use for transcriptome assembly, along with the operable test of the interview.

VITA

Adam M. Rork

EDUCATION

The Pennsylvania State University, University Park, PA — 2017-2021
Ph.D., Entomology

The Pennsylvania State University, University Park, PA — 2018-2019
Graduate Certificate, Applied Bioinformatics

Maryville University of St. Louis, St. Louis, MO — 2013-2017
B.S., Biochemistry

RESEARCH EXPERIENCE

Graduate Research Fellow/Assistant — 2017-2021
The Pennsylvania State University

NSF-REU Intern — Summer 2015
The Chicago Botanic Garden

Undergraduate Research Assistant — 2014-2017
Maryville University/Missouri Botanical Garden

PUBLICATIONS

Rork AM, Xu S, Attygalle A, Renner T. (2021). Primary Metabolism Co-opted for Defensive Chemical Production in the Carabid Beetle, *Harpalus pensylvanicus*. Journal of Chemical Ecology. 47(3): 334-349.

Rork AM, Mikó I, Renner T. (2019). Pygidial Glands of *Harpalus pensylvanicus* (Coleoptera: Carabidae) Contain Resilin-rich Structures. Arthropod Structure & Development. 44(12): 1069-1083.

Rork AM, Renner T. (2018). Carabidae Semiochemistry: Current and Future Directions. Journal of Chemical Ecology. 49: 19-25.

GRANTS & AWARDS

Michael E. Duke Memorial Scholarship in Entomology — 2021
The Pennsylvania State University

Postdoctoral Research Fellowship Biology — 2022
The National Science Foundation

Graduate Student Research Enhancement Award — 2020
The Coleopterists Society

TEACHING EXPERIENCE & OUTREACH

Graduate Teaching Assistant (ENT 202, 313, 432) — 2017-2018

Great Insect Fair Participant — 2017-2019

CSTAS Continuing Education Workshop Participant — 2018

ENVISION Workshop Participant — 2019-2020