Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Wen-Hao Tan Date

Ecological immunology in a tri-trophic context: plant-mediated defense and immune gene evolution in monarch butterflies

By

Wen-Hao Tan
Doctor of Philosophy

Graduate Division of Biology and Biomedical Sciences
Population Biology, Ecology, and Evolution

_____
Jacobus C. de Roode, Ph.D.
Advisor

_____
Nicole M. Gerardo, Ph.D.
Advisor

_____
Christopher W. Beck, Ph.D.
Committee Member

_____
Levi T. Morran, Ph.D.
Committee Member

_____
Michael R. Strand, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Ecological immunology in a tri-trophic context: plant-mediated defense and immune gene evolution in monarch butterflies

By

Wen-Hao Tan
B.S. National Taiwan University, 2013

Advisor: Jacobus C. de Roode, Ph.D.
Advisor: Nicole M. Gerardo, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Graduate Division of Biological and Biomedical Sciences
Population Biology, Ecology, and Evolution
2019

Abstract

Ecological immunology in a tri-trophic context: plant-mediated defense and immune gene evolution in monarch butterflies

By Wen-Hao Tan

Animals have evolved multiple forms of defense against parasites. Animal defense is often studied in the context of pairwise interactions between hosts and parasites, but the environmental context in which they interact can also shape animal defense. For instance, toxic plant secondary chemicals can increase herbivore defense against their enemies. Thus, specialization on toxin-producing plants may shape the evolution of other herbivore defenses, leading to variation within and between populations. In this dissertation, I used monarch butterflies (*Danaus plexippus*) to examine plant-mediated defense and immune gene evolution in a tri-trophic context involving the butterflies, their parasites (*Ophryocystis elektroscirrha*), and their milkweed hostplants (*Asclepias* spp.). In this system, milkweed species affects monarch resistance to the parasites, and this difference in resistance is correlated with plant toxicity. First, I studied how milkweed inducible toxin production upon herbivory affects monarch parasite resistance. I found that interspecific variation in plant toxicity is a more important driver of parasite resistance than plasticity in toxin production via induction. Second, I studied changes in global gene expression of monarchs in response to milkweed diet and parasite infection, and the interplay between plant toxins and herbivore immunity. I found that monarchs differentially express several hundred genes when feeding on plant species that differ in toxicity, including genes belonging to multiple families of canonical detoxification genes, which may play a role in toxin resistance and sequestration. Also, I found that a small number of immune genes were down-regulated when feeding on a more toxic plant. Third, I studied the evolution of canonical immune genes across different monarch populations that vary in their association with milkweed species and parasite prevalence. I found that different classes of immune genes exhibit different patterns of selection, differentiation, and polymorphism. For example, signaling genes exhibit signals of purifying selection while effector genes show signatures of balancing selection. I also found some population-specific patterns, suggesting that monarch immune genes are not under uniform selection across geographically, genetically and ecologically distinct populations. By combining experimental approaches, studies of gene expression, and population genomic analyses, this study increases our understanding of the evolutionary ecology of herbivore immunity.

Ecological immunology in a tri-trophic context: plant-mediated defense and immune gene evolution in monarch butterflies

By

Wen-Hao Tan
B.S. National Taiwan University, 2013

Advisor: Jacobus C. de Roode, Ph.D.
Advisor: Nicole M. Gerardo, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
Graduate Division of Biological and Biomedical Sciences
Population Biology, Ecology, and Evolution
2019

Acknowledgments

I would like to thank my advisors, Nicole Gerardo and Jaap de Roode, for being amazing mentors. They contributed their time, knowledge, and resources to guide and support my research. With their invaluable advice, I am able to use multiple approaches, including greenhouse experiments, molecular and immunological experiments, genomic and bioinformatic analyses, to study on a system and research topics that I am deeply interested in from multiple aspects. I would like to thank my dissertation committee, Chris Beck, Levi Morran, and Mike Strand (U. Georgia) for their guidance throughout the process, and Gonzalo Vazquez-Prokopec and Lance Waller for serving on my qualifying exam committee. They challenged me to broaden my knowledge in the field and their insightful comments greatly improved this dissertation work.

Outside of my committee, I would like to thank Jamie Walters (U. Kansas) for teaching me bioinformatics and population genomics, and for his advice and input on two of my chapters. I am grateful for our tight remote collaboration, which he contributed his time and expertise, and generously provided me access to their powerful computing cluster at U. Kansas, making the bioinformatics component of my dissertation possible. I also thank Andrew Mongue from the Walters lab for collaborating on chapter four. In addition, I would like to thank Mark Hunter (U. Michigan) for his advice and input on two of my chapters, and his technician for performing all the chemical analyses.

I would like to thank all members of the Gerardo and de Roode labs for creating a fun and supportive working environment, including Leiling Tao, Erika Diaz-Almeyda, Aileen Berasategui, Hassan Salem, Miguel Reyes, Kandis Adams, Caitlin Conn, Scott Villa, Venkat Talla, Lewis Bartlett, Tarik Acevedo, Tiffani Alcaide, Joselyne Chavez, Erik Edwards, Amanda Pierce, Justine Garcia, Zachary Lynch, Mary Bushman, Erica Harris, Travis Dynes, Kim Hoang, Signe White,

Sandra Mendiola, Kayla Stoy, Kevin Hoang, and all the undergraduate students in both labs. Especially, I would like to thank Leiling for mentoring me during my rotation in the de Roode lab, training me in monarch rearing and greenhouse maintenance, and collaborating on chapter one. I thank Tarik, Erica, Tiff, and Kevin for collaborating on experiments, Hassan, Aileen, Erika, Amanda, and Venkat for providing substantial feedback to my projects and manuscripts. Also, I am thankful to all the undergraduate and high school students that I had worked with, including Yaw Kumi-Ansu, Jessie Cai, Stacey Kim, Kristen Echimer, Becca Nation, and Alicia Johnson.

I would like to thank the entire PBEE community for their support. Especially, I thank my PBEE cohort Kim Hoang, Signe White, Kathryn Schaber, Mari Peterson for their support and going through qualifying exams together, and Robert Petit for teaching me Python. Also, I thank Berry Brosi for offering an opportunity to do a rotation in his lab.

Neither pursuing a Ph.D. nor studying abroad are simple challenges. I thank my biology teacher and advisors during high school and college, Tsui-Hua Liu, Yu-Feng Hsu, and Yu-Teh Kirk Lin, for inspiring my interest in science, teaching me how to perform scientific research, and preparing me for the challenges in graduate school and studying abroad. Also, I thank all the support from my friends back in Taiwan and here in Atlanta.

Finally, I would like to thank my parents for their unconditional support throughout. It was a tough challenge being very far away from them and the rest of my family, and I would not have been able to do it without their support and encouragement. Most importantly, I would like to thank my wife, Yu-Hui Lin, who just got her Ph.D. degree too. We went through all the ups and downs together during this journey, including moving from Taiwan to the United States, everything happening in the last five years in Atlanta, and now moving back to Taiwan. Her support and encouragement are always critically important to me.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Immunity in a tri-trophic context

Parasites are a major threat to their host's fitness, and natural selection should favor defense mechanisms that can protect hosts against parasites. Consequently, hosts have evolved multiple defenses, including mechanisms to prevent infection, reduce parasite growth, and alleviate disease symptoms (Parker, Barribeau, Laughton, de Roode, & Gerardo, 2011). These defenses, such as the cellular and humoral immune system, are under strong selection pressures exerted by parasites (McTaggart, Obbard, Conlon, & Little, 2012; Schlenke & Begun, 2003). Although most studies on host defense against parasites have focused on immune system-mediated defenses (Little, Hultmark, & Read, 2005), hosts can also use alternative mechanisms to defend against parasites (Parker et al., 2011). Host defenses against parasites are often studied within a framework of isolated pair-wise interactions; however, the environment in which they are embedded can influence their interactions. This environmental influence is rarely considered in host-parasite studies (Lazzaro & Little, 2009; Wolinska & King, 2009). As suggested by the field of ecological immunology, environmental context and long-term associations with important ecological factors could shape the evolution of animal immunity, and maintain variation within and between populations. In this dissertation, I studied insect immunity in a plant-herbivore-parasite tri-trophic context.

1.2 Plant-mediated effects on herbivore defenses

Plants have evolved multiple forms of defense, including the production of toxic secondary

metabolites to protect themselves from herbivory (Schoonhoven, van Loon, & Dicke, 2005). Herbivorous insects have evolved mechanisms to overcome such plant defenses, including contact avoidance, rapid excretion, sequestration, enzymatic detoxification, and target site mutation (Després, David, & Gallet, 2007). A suite of physiological changes may be required to cope with multiple forms of plant defense. Further, herbivorous insects that can feed on multiple plant species have the additional complication that they may encounter different quantities and/or classes of phytochemicals when feeding on different host plants. In addition, some plants can increase their production of secondary chemicals upon herbivory, a phenomenon known as phytochemical induction (Karban & Baldwin, 1997). In such a scenario, a plastic response of herbivore defense may be favored when herbivorous insects are associated with plant diets with varying levels of toxicity.

Phytochemicals are an important ecological context that can mediate multi-trophic interactions. For instance, some specialist insects are able to sequester plant secondary metabolites into their own tissues, thereby protecting themselves against their own natural enemies (Nishida, 2002; Opitz & Müller, 2009). Through trophic interactions, plant defenses can result in either net positive or net negative effects on plant fitness (Cory & Hoover, 2006). Therefore, when specialist herbivores sequester plant toxins and use them for their own defense, defenses induced in the plant in response to herbivory present a problem for plants as they confer benefits to their herbivores (Fordyce, 2001). This question forms the basis of chapter 2, in which I studied how plant inducible defense influences herbivore resistance against parasites.

Differences in chemical composition between plant species, such as nutritional properties or toxic chemical contents, can alter herbivore disease resistance. Immune systems often require

abundant nutritional resources and energy to maintain proper function (Glick, Day, & Thompson, 1981; Glick et al., 1983). For example, the immune responses of *Spodoptera littoralis*, a herbivorous insect, are significantly stronger when fed with high-quality protein diet (Lee, Simpson, & Wilson, 2008). Many plants produce toxic secondary metabolites as a defense against herbivores, and those toxins may impair immune function. Alternatively, secondary chemicals can alter herbivore disease resistance by conferring direct negative effects on parasites and pathogens or by altering host immune function (Lampert, 2012; Smilanich, Dyer, Chambers, & Bowers, 2009). For herbivores that specialize on toxin-producing plants, previous studies have demonstrated that plant diets with high toxicity can either reduce or enhance their immune responses. For instance, feeding on plants with high levels of iridoid glycosides (IGs) enhance encapsulation responses in the Glanville fritillary (*Melitaea cinxia*) (Laurentz et al., 2012); however, feeding on high-IG plants reduce encapsulation responses in *Junonia coenia* (Smilanich et al., 2009). In addition, plant diet can also alter or mediate alternative defense mechanisms, such as behavioral defenses, especially when the diet contains chemicals that directly confer anti-parasitic effects. For instance, some animals can self-medicate by ingesting toxic plant chemicals when infected by parasites or pathogens (Abbott, 2014; de Roode, Lefèvre, & Hunter, 2013; Singer, Mace, & Bernays, 2009). Altogether, plant secondary chemicals can modulate multiple forms of herbivore defense against their natural enemies.

An intuitive expectation is that animals will maximize their investment in multiple forms of defense at once to gain the strongest protection against parasites. However, the field of ecological immunology suggests that costly defenses exhibit trade-offs with other life-history traits (Sheldon & Verhulst, 1996). When plant chemicals strengthen one form of defense, they could

simultaneously lessen investment in other defense mechanisms, because costly biological processes often exhibit trade-offs due to physiological constraints (Ricklefs & Wikelski, 2002; Stearns, 1989). That is, an animal may not be able to increase investment in immunity without decreasing investment in other traits due to energy allocation. Similarly, detoxification and sequestration of plant toxins can be energetically costly (Bowers, 1992), so a reduction in immune function could be caused by trade-offs with these processes (Moret & Schmid-Hempel, 2000). For example, studies on IG-sequestering herbivorous insects have demonstrated that an increased level of IG sequestration correlates with a decreased level of immune responses, presumably due to energetic costs (Lampert & Bowers, 2015; Quintero, Lampert, & Bowers, 2014; Richards et al., 2012; Smilanich et al., 2009). Alternatively, an animal may not need to invest in multiple defenses if one is sufficient, and thus costs can be reduced by utilizing one defense in place of others. Honey bees (*Apis mellifera*) and wood ants (*Formica paralugubris*) that behaviorally incorporate anti-microbial resins into their nests have reduced investment in immune gene expression and humoral antimicrobial activity (Castella, Chapuisat, Moret, & Christe, 2008; Simone, Evans, & Spivak, 2009). These results suggest that the use of medicinal plants may make immune system-mediated defenses superfluous.

Taken together, plant secondary chemicals can play an important role in modulating herbivore defenses, including conferring direct negative effects on parasites and pathogens, enhancing immune system functions, or reducing immune investment. This question leads to the study of chapter 3, in which I carried out a transcriptomic analysis to examine the interplay between plant toxins and parasite infection on global gene expression patterns of a herbivorous insect.

1.3 Evolution of herbivore immunity in association with toxic phytochemicals

Long-term associations with particular plant secondary chemicals could shape the evolution of herbivore defenses (Smilanich, Dyer, & Gentry, 2014). Adaptation to toxic plants may either decrease or increase the evolutionary maintenance of immune response. If medicinal chemicals confer direct anti-parasitic effects, it could make immune response superfluous, relaxing selection to maintain immune function. In contrast, if medicinal chemicals stimulate immunity, it could strengthen reliance on immune responses, exerting selection to maintain immune system function. In many systems, there is considerable intra- and inter-specific variation in toxicity of plants, which could lead to either increased specialization or plasticity to cope with this variable environmental context. Altogether, the use of medicinal plants could shape the evolutionary maintenance of defense, diet choice, and plasticity, and populations may vary in immune functions based on their ecological context.

Diet-mediated effects on defenses have been considered one of the major factors shaping the evolution of diet choice and diet breadth in herbivorous insects (Ojala, Julkunen-Tiitto, Lindström, & Mappes, 2005; Smilanich et al., 2009). Specialist species and generalist species can differ in their response to plant chemicals, and can exhibit differences in the way those chemicals directly or indirectly mediate defenses. *Ceratomia catalpae*, a moth species that specializes on host plants with high toxin levels, for example, has better sequestration efficacy than the generalist *C. undulosa*, thus having stronger chemical defenses (Lampert & Bowers, 2015). The availability of a diverse set of dietary resources is important for some generalist insects to maintain strong immune responses (Muller, Vogelweith, Thiéry, Moret, & Moreau, 2015), and the benefit of having better constitutive levels of immune system-mediated defenses can potentially have fitness advantages

that outweigh the costs of finding and metabolizing a diverse diet. Some studies have shown, for example, that polyphagy strongly enhances immune system-mediated defenses in European grapevine moths (*Lobesia botrana*) compared to feeding on a single food source (Muller et al., 2015). Taken together, within-population selection on diet choice can sometimes be driven by maintaining defenses, highlighting the importance of considering tri-trophic interactions when studying the evolution of herbivore defenses.

Herbivore-plant associations, specifically the availability of medicinal hostplants, may vary across populations. For specialist herbivores, in a populations with access to diets that stimulate defenses, selection can favor individuals with greater ability to utilize these diets to maximize defense strength, thus influencing the evolution of diet choice. Furthermore, the availability of medicinal resources can also favor the use of different defense mechanisms under different environmental contexts. Therefore, populations or closely related species associated with different resources may have evolved differences in defense traits (Lampert & Bowers, 2015), leading to population differentiation and local adaptation. Alternatively, under the same premise that the availability of medicinal chemicals varies between natural habitats, varying selection pressures could favor the evolution of phenotypic plasticity rather than divergence (Lande, 2009; Miner, Sultan, Morgan, Padilla, & Relyea, 2005; Torres-Dowdall, Handelsman, Reznick, & Ghalambor, 2012). This question forms the basis of chapter 4, in which I carried out a population genomic study to examine immune gene evolution across natural populations varying in their environmental context.

It has been hypothesized that the use of alternative defense mechanism may make cellular and humoral immunity superfluous, relaxing selection to maintain these responses (de Roode et al., 2013; Sadd & Schmid-hempel, 2009; Sheldon & Verhulst, 1996). However, there are few empirical

studies demonstrating evolutionary trade-offs between alternative defenses at a microevolutionary scale (within species). Given both intra- and inter-specific variation of plant toxicity and differences in host plant community between natural habitats, we can hypothesize that populations or closely related species adapted to different host plant species across habitats may show evolutionary differences in the strength of defense traits if medicinal chemicals influence one or more of those defenses. In general, few studies have examined the relationship between the strength of multiple defense mechanisms across populations, and even less is known regarding whether this variation is tied to the availability of medicinal compounds. There is, however, some indirect evidence in insect systems. For instance, honey bees (*Apis mellifera*) have a variety of behavioral defense mechanisms such a self-medication, and analysis of their genomes has shown a reduction in the number of canonical insect immune genes compared to fruit flies and mosquitoes (Evans et al., 2006; Simone et al., 2009). Pea aphids (*Acyrthosiphon pisum*) harbor beneficial symbiotic bacteria that confer protection against parasitoids and pathogens, and analyses of their genome also indicate that they lack many insect immune genes (Gerardo et al., 2010; Oliver, Russell, Moran, & Hunter, 2003; Parker, Spragg, Altincicek, & Gerardo, 2013).

1.4 Monarch butterflies as a study system

Monarch butterflies (*Danaus plexippus*) are specialist herbivores on milkweed plants (mostly *Asclepias* spp.) and are distributed worldwide in several genetically distinct populations (Ackery & Vane-Wright, 1984; Pierce et al., 2014). Monarchs worldwide are commonly infected with a protozoan parasite, *Ophryocystis elektroscirrha* (Altizer, Oberhauser, & Brower, 2000; Leong,

Yoshimura, & Kaya, 1997), making them a experimentally tractable natural system for studying infectious disease ecology and evolution. *O. elektroscirrha* prevalence differs greatly between monarch populations, ranging from 8% (*e.g.*, North America population) to over 70% highly infected (*e.g.*, south Florida population) (Altizer & de Roode, 2015; Altizer et al., 2000). *O. elektroscirrha* spores transmit vertically from female monarchs to their offspring during oviposition on milkweeds, and can be transmitted horizontally as well. Following ingestion of parasite spores by larvae, sporozoites are released from the spores, and then penetrate the gut wall and replicate in hypodermal tissues. During the host's pupal stage, they replicate sexually and form spores in the tissues that are destined to develop into adult scales, which enables the parasite to transmit externally during oviposition (Fig. 1.1) (Mclaughlin & Myers, 1970). *O. elektroscirrha* infection reduces host fitness, and the reduction in fitness correlates positively with parasite spore loads (de Roode, Yates, Altizer, & Roode, 2008). Disease symptoms are expressed during the adult stage and include reduced body mass, mating ability, and lifespan (de Roode, Yates, et al., 2008).

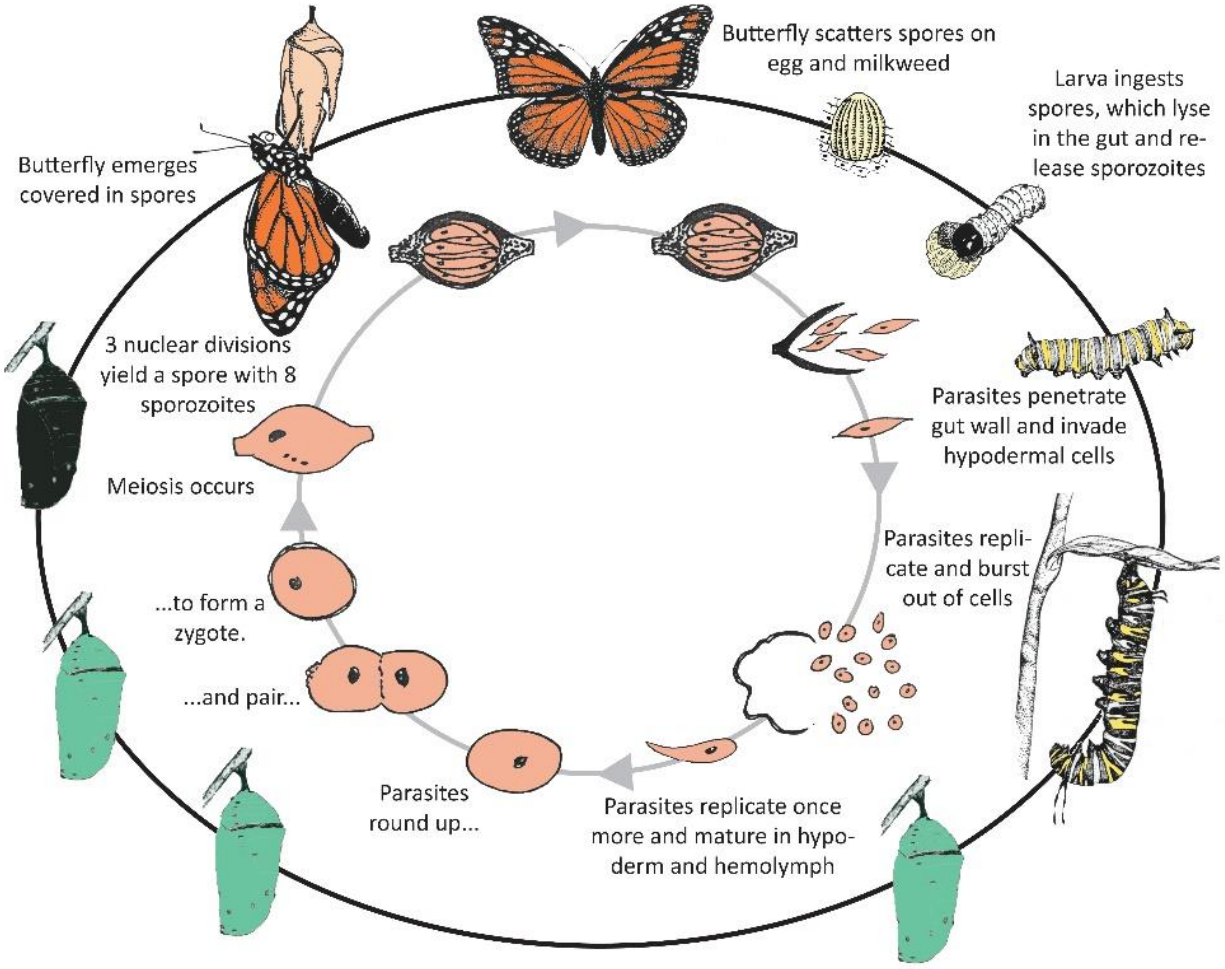


Figure 1.1. The life cycles of monarch butterflies (outside circle) and *O. elektroscirrha* parasites (inside circle). Figure credit: Jacobus de Roode.

Monarch butterflies are a prominent example of sequestration and aposematism, as they use milkweed toxins to deter bird predators (Agrawal, Petschenka, Bingham, Weber, & Rasmann, 2012; Brower, Ryerson, Coppinger, & Susan, 1968). In addition to the anti-predator protection, previous studies have shown that some milkweeds can reduce *O. elektroscirrha* infection and spore load (Sternberg et al., 2012; Tao, Hoang, Hunter, & de Roode, 2016). Different milkweed species vary greatly in their anti-parasitic properties (Fig. 1.2), with higher concentrations of cardenolides, toxic

plant secondary chemicals, confers greater resistance against parasite infection (Sternberg et al., 2012). Consumption of milkweeds with high levels of cardenolides in the larval stage before and during infection correlates with reduced disease symptoms (Gowler, Leon, Hunter, & de Roode, 2015; Sternberg et al., 2012). On the other hand, feeding on milkweeds with high levels of cardenolides also incurs fitness costs, including reduced larval survival and adult lifespan (Agrawal, 2005; Malcolm, 1994; Tao, Hoang, et al., 2016; Zalucki & Brower, 1992; Zalucki, Brower, & Alonso-M, 2001; Zalucki, Brower, & Malcolm, 1990). High-cardenolide milkweeds may confer protection through direct toxic effects to the parasite or through indirect alteration of host immunity, but the mechanism remains unknown. In addition, infected adults show trans-generational medicating behavior - when given a dual choice between high- and low-cardenolide milkweeds (*A. curassavica* and *A. incarnata*) in the laboratory, infected adults preferentially lay eggs on the high-cardenolide milkweed, thereby increasing their offspring's resistance to the parasite (Lefèvre et al., 2012; Lefèvre, Oliver, Hunter, & de Roode, 2010).

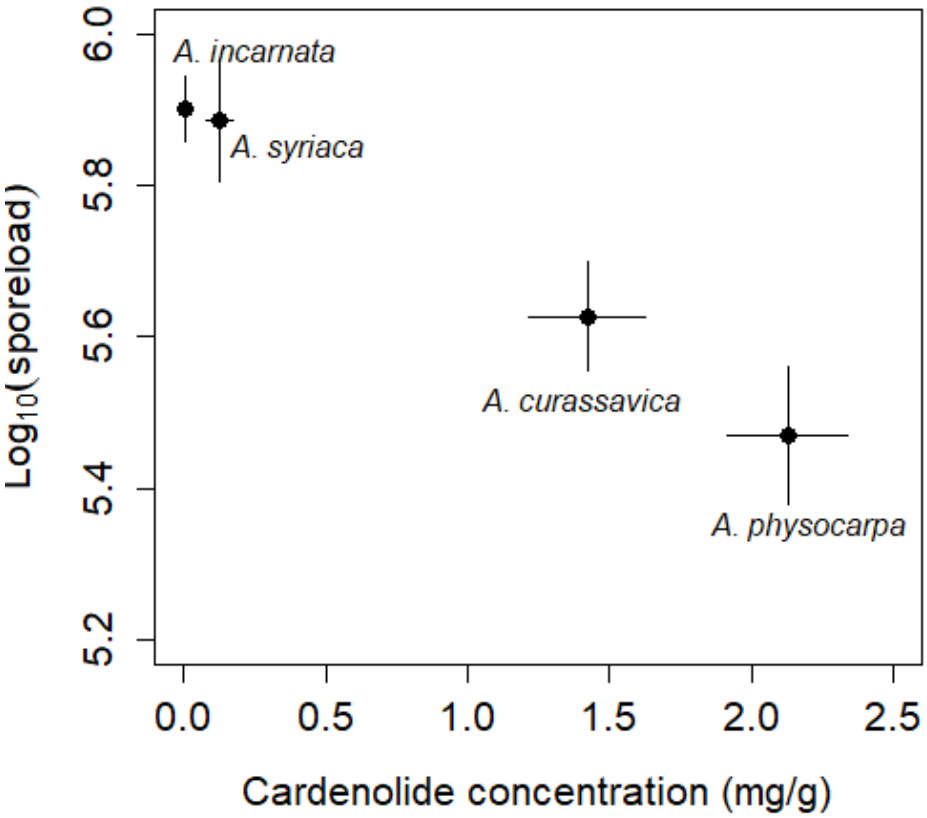


Figure 1.2. Anti-parasitic effects of four milkweed species varying in toxicity. This graph includes all milkweed species studied in this dissertation (Tan *et. al.* unpublished).

Monarchs have a wide distribution, and different geographic populations are associated with different environmental contexts. Monarchs originated in North America and colonized worldwide locations in the 19th century through independent dispersal events across the Pacific Ocean, the Atlantic Ocean, and across Central-South American (Fig. 1.3) (Ackery & Vane-Wright, 1984; Pierce et al., 2014; Zhan et al., 2014). During the dispersal process, monarchs formed populations in which they relied on more toxic milkweed host plants and in which they experienced greater risk of infection by the specialist parasite *O. elektroscirrha* (Altizer & de Roode, 2015). How these important ecological factors and dispersal events may shape the evolution of monarch immunity

remains unknown.

Figure 1.3. Current distribution of monarch butterflies around the world and their historical dispersal routes. Monarchs originated in the North America and established other populations via three main dispersal events: across the Pacific Ocean, across the Atlantic Ocean, and toward Central/South America. Figure credit: Jacobus de Roode.

In this thesis, I used the milkweed-monarch-parasite system (Fig. 1.4) to study plant-mediated defense and immune gene evolution in monarch butterflies. In chapter 2, I studied how inducible defenses of milkweeds affect parasite resistance of monarchs. In chapter 3, I studied the global gene expression of monarch larvae in response to milkweed diet and parasite infection, and the potential interactions between the two. In chapter 4, I studied the evolution of canonical immune genes across different monarch populations that vary in with their association with milkweed species and parasite prevalence.

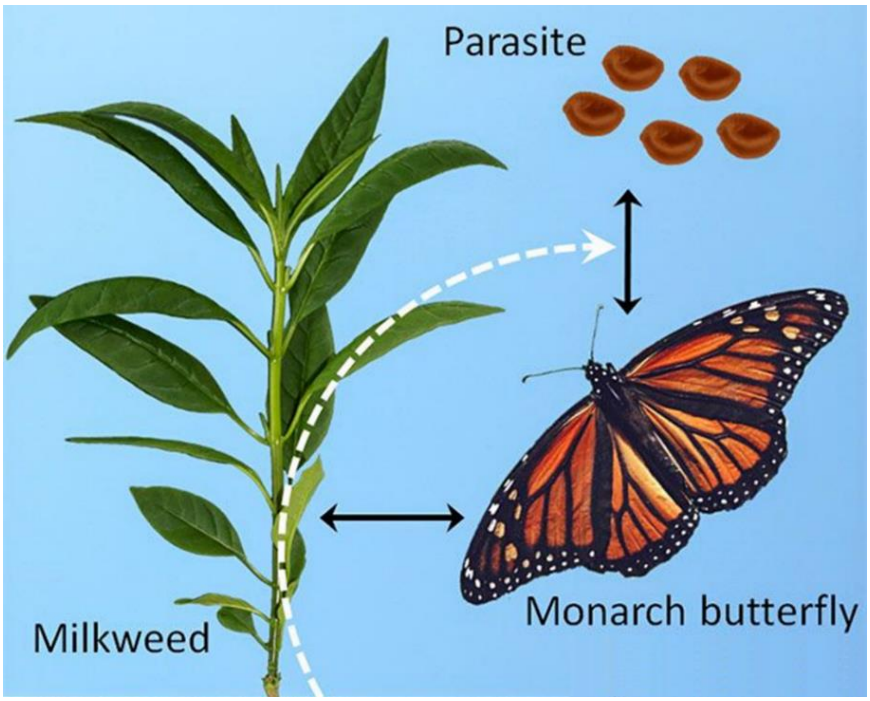


Figure. 1.4. Tri-trophic interactions between milkweeds, monarchs, and parasites. Figure credit: Jacobus de Roode.

Chapter 2: The effects of milkweed induced defense on parasite resistance in monarch butterflies, *Danaus plexippus*

Reprinted material from: Wen-Hao Tan, Leiling Tao, Kevin M. Hoang, Mark D. Hunter, and Jacobus C. de Roode (2018) The effects of milkweed induced defense on parasite resistance in Monarch butterflies, *Danaus plexippus*. *Journal of Chemical Ecology* 44:1040–1044.

Abstract

Many plants express induced defenses against herbivores through increasing the production of toxic secondary chemicals following damage. Phytochemical induction can directly or indirectly affect other organisms within the community. In tri-trophic systems, increased concentrations of plant toxins could be detrimental to plants if herbivores can sequester these toxins as protective chemicals for themselves. Thus, through trophic interactions, induction can lead to either positive or negative effects on plant fitness. We examined the effects of milkweed (*Asclepias* spp.) induced defenses on the resistance of monarch caterpillars (*Danaus plexippus*) to a protozoan parasite (*Ophryocystis elektroscirrha*). Milkweeds contain toxic secondary chemicals called cardenolides, higher concentrations of which are associated with reduced parasite growth. Previous work showed that declines in foliar cardenolides caused by aphid attack render monarch caterpillars more susceptible to infection. Here, we ask whether cardenolide induction by monarchs increases monarch resistance to disease. We subjected the high-cardenolide milkweed *A. curassavica* and the low-cardenolide *A. syriaca* to caterpillar grazing, and reared infected and uninfected caterpillars

on these plants. As expected, monarchs suffered less parasite growth and disease when reared on *A. curassavica* than on *A. syriaca*. We also found that herbivory increased cardenolide concentrations in *A. curassavica*, but not *A. syriaca*. However, cardenolide induction in *A. curassavica* was not strong enough to influence monarch resistance to the parasite. Our results suggest that interspecific variation in cardenolide concentration is a more important driver of parasite defense than plasticity via induced defenses in this tri-trophic system.

2.1 Introduction

Plants have evolved multiple forms of defense against herbivores, including the production of secondary chemicals that are toxic to animals. In addition to producing constitutive levels of toxic compounds, some plants increase their production of secondary chemicals upon herbivory, a phenomenon known as phytochemical induction (Karban & Baldwin, 1997). This increase in toxic phytochemicals can have direct or indirect effects on other trophic levels, with either net positive or net negative effects on plant fitness (Cory & Hoover, 2006). Specifically, if plant induced defenses have negative effects on parasites or pathogens of herbivores, they indirectly lower plant fitness through trophic interactions (Hunter & Schultz, 1993). When specialist herbivores sequester plant toxins and use them for their own defense, induced defenses present a problem for plants as they confer benefits to their herbivores (Fordyce, 2001).

In this study, we examined the effects of phytochemical induction on higher-trophic interactions, specifically herbivore-parasite interactions, with a herbivore that sequesters plant toxins as protective chemicals. We studied milkweeds (*Asclepias* spp.), monarch butterflies (*Danaus plexippus*), and a specialist protozoan parasite (*Ophryocystis elektroscirrha*). Milkweeds

contain toxic cardenolides that disrupt animal Na^+/K^+-ATPase (Agrawal, Petschenka, et al., 2012), and monarch caterpillars sequester cardenolides as protective chemicals against bird predators (Brower, McEvoy, Williamson, & Flannery, 1972). *O. elektroscirrha* spores are ingested by monarch caterpillars as they feed on milkweed and can be transmitted vertically from females to their offspring during oviposition. Parasite infection reduces host fitness, by lowering pre-adult survival, mating ability, and adult lifespan (de Roode, Yates, et al., 2008). A body of work has shown that the consumption of milkweed species with higher concentrations of cardenolides also increases monarch resistance to infection with the parasite (*e.g.*, Sternberg et al., 2012). Moreover, declines in foliar cardenolide concentrations resulting from aphid attack render monarch caterpillars more susceptible to the parasite (de Roode, Rarick, Mongue, Gerardo, & Hunter, 2011). In contrast with aphid-associated reductions in cardenolides, some milkweed species increase their concentrations of cardenolides following herbivory by monarchs (Agrawal, Hastings, Patrick, & Knight, 2014; Agrawal, Kearney, Hastings, & Ramsey, 2012). Thus, we hypothesized that the induced upregulation of cardenolides by herbivory could enhance monarch resistance to the parasite.

2.2 Methods and Materials

2.2.1 Plants, Butterflies, and Parasites.

Three species of milkweed were used for the experiment: *A. curassavica, A. syriaca,* and *A. incarnata*. Milkweed seeds were obtained from Prairie Moon Nursery (Winona, MN, USA). All milkweeds used for this study were about 3 months old, grown in a greenhouse under natural light conditions with weekly fertilization. Monarchs were obtained from five lab-reared outcrossed

lineages generated from wild-caught migratory monarchs collected in St. Marks, Florida, USA. The parasite clone used was generated from an infected, wild-caught monarch from the same population.

2.2.2 Experimental Design and Procedures.

Our experiment represented a 2 x 2 x 2 fully factorial design, with plant species, induction, and parasite infection as main factors. Two milkweed species, *A. curassavica* and *A. syriaca*, were used for the induction treatments. *A. curassavica* leaves contain higher cardenolide concentrations than do those of *A. syriaca*, and monarchs reared on *A. curassavica* experience reduced parasite growth and disease symptoms (*e.g.*, Sternberg et al., 2012). For each species, plants were divided between induced and control treatments; within each induction/control treatment, plants were further divided between infected and uninfected treatments. Individual plants in the induced treatment were treated with one uninfected second instar larva, which was allowed to feed for 5 days, a timeframe that is sufficient to trigger induced defenses in milkweed (Agrawal et al., 2014). The herbivory treatment caused about 20-50% of leaf damage by visual assessment. Control uninduced plants did not receive grazing caterpillars. After 5 days of induction treatment, an 8-mm diameter leaf disk was taken from each experimental plant for parasite inoculation. Second instar monarch larvae were inoculated by adding 10 parasite spores to the leaf disk taken from their pre-assigned plant (de Roode, Rarick, et al., 2011; Sternberg et al., 2012). Uninfected controls received disks without spores. Before inoculation, all larvae had been reared on *A. incarnata*, a low cardenolide milkweed species, to homogenize their dietary experience prior to treatment. After larvae consumed their entire leaf disk, and therefore parasite dose, larvae were transferred to individual rearing cups (473 mL) and fed with leaf cuttings from the same individual plant that was

used for parasite inoculation. Thus, in our experiment, phytochemical induction occurred on live plants, which were then subsampled to feed experimental caterpillars. Previous work (de Roode, Fernandez de, Faits, & Alizon, 2011) has shown that the effects of cardenolides on parasite infection are mediated during the time of infection, and that the milkweed fed to caterpillars fed following infection has no further effects on parasitism. Thus, feeding caterpillars with leaf disks from induced and uninduced plants, followed by the feeding of leaf cuttings, provides a relevant approach to study the effects of cardenolide induction on monarch susceptibility to parasites. Sample sizes ranged from 13 – 25 per treatment group. After pupation, pupae were placed in a laboratory room maintained at 25 °C under 14/10h L/D cycle. After eclosion, adults were placed in 8.9 x 8.9 cm glassine envelopes without a food source at 12 °C under 14/10h L/D cycle. Previous studies have demonstrated that the effects of parasite infection on adult longevity are similar between starvation conditions and more natural, non-starvation conditions (de Roode, Chi, Rarick, & Altizer, 2009). Adults were inspected daily until death to measure lifespan. Parasite load was quantified using a vortexing protocol described in de Roode *et al*. (2011).

2.2.3 Chemical Analyses.

Plant chemical samples were collected on the same day as parasite inoculation, which was after 5 days of experimental herbivory. One leaf from the fourth pair on each plant was chosen, and six leaf disks (424 mm^2 total) were taken with a paper hole punch from one side of the leaf and placed immediately into a 1 mL collection tube with cold methanol. Another six identical leaf disks were taken from the opposite side of the same leaf to measure sample dry mass. Total cardenolide concentrations were analyzed using reverse-phase ultra-performance liquid chromatography (UPLC; Waters Inc., Milford, MA, USA) following established methods (de Roode, Rarick, et al.,

2011; Tao, Gowler, Ahmad, Hunter, & de Roode, 2015).

2.2.4 Statistical Analyses.

The main goal of our analyses was to test for the effects of plant species and induction on foliar cardenolide concentration, parasite resistance, and adult lifespan (see below for definition for each). We tested the main effects of plant species, induction, and/or infection as fixed effects using linear models. Normality and variance homogeneity were checked with the Shapiro-Wilk normality test and Fligner-Killeen test. Cardenolide concentration was analyzed separately by plant species using a Mann-Whitney test with Bonferroni corrections for multiple comparisons due to violation of assumptions of normality and variance homogeneity. Parasite resistance data were Box-Cox transformed and analyzed using linear models with weighted least squares to meet model assumptions, with parasite load as dependent variable and plant species, induction, and their interaction as independent variables. Lifespan data were Box-Cox transformed to meet assumptions and analyzed separated by species using linear models with Bonferroni corrections for multiple comparisons with lifespan as dependent variable and infection, induction, and their interaction as independent variables. *Post hoc* Tukey pairwise comparisons were performed when the interaction term in a linear model was significant. All analyses were performed using R version 3.4.1. Box-Cox transformation was performed with the package car 2.1-5. *Post hoc* Tukey pairwise comparisons were performed with the package multcomp 1.4-7.

2.3 Results

Herbivory by monarch caterpillars induced significantly higher cardenolide concentrations in

A. curassavica foliage (Fig. 2.1A; $W = 398$, $P < 0.001$), but not in *A. syriaca* foliage (Fig. 2.1A; $W = 514$, $P = 0.96$). *A. curassavica* foliage also had higher constitutive cardenolide concentrations (Fig. 2.1A). Monarchs reared on *A. curassavica* had significantly lower spore loads than monarchs reared on *A. syriaca*, but induction did not significantly affect spore load (Fig. 2.1B; plant species: $F_{1, 80} = 8.55$, $P < 0.01$; induction: $F_{1, 80} = 0.09$, $P = 0.76$; plant species x induction: $F_{1, 80} = 1.55$, $P = 0.22$).

Although induction did not affect parasite spore load, it reduced the lifespan of some monarchs. Specifically, uninfected larvae reared on induced *A. syriaca* lived shorter lives than did uninfected larvae reared on control *A. syriaca* (Fig. 2.1C; infection: $F_{1, 60} = 571.50$, $P < 0.01$; infection x induction: $F_{1, 60} = 6.67$, $P = 0.01$). Induction had no such effects on monarchs reared on *A. curassavica*, although infection reduced lifespan as expected (Fig. 2.1C; infection: $F_{1, 72} = 246.45$, $P < 0.01$; infection x induction: $F_{1, 72} = 3.64$, $P = 0.06$). Overall, our results suggest that induced cardenolide defenses of milkweeds have weak effects on monarch defenses against the parasite.

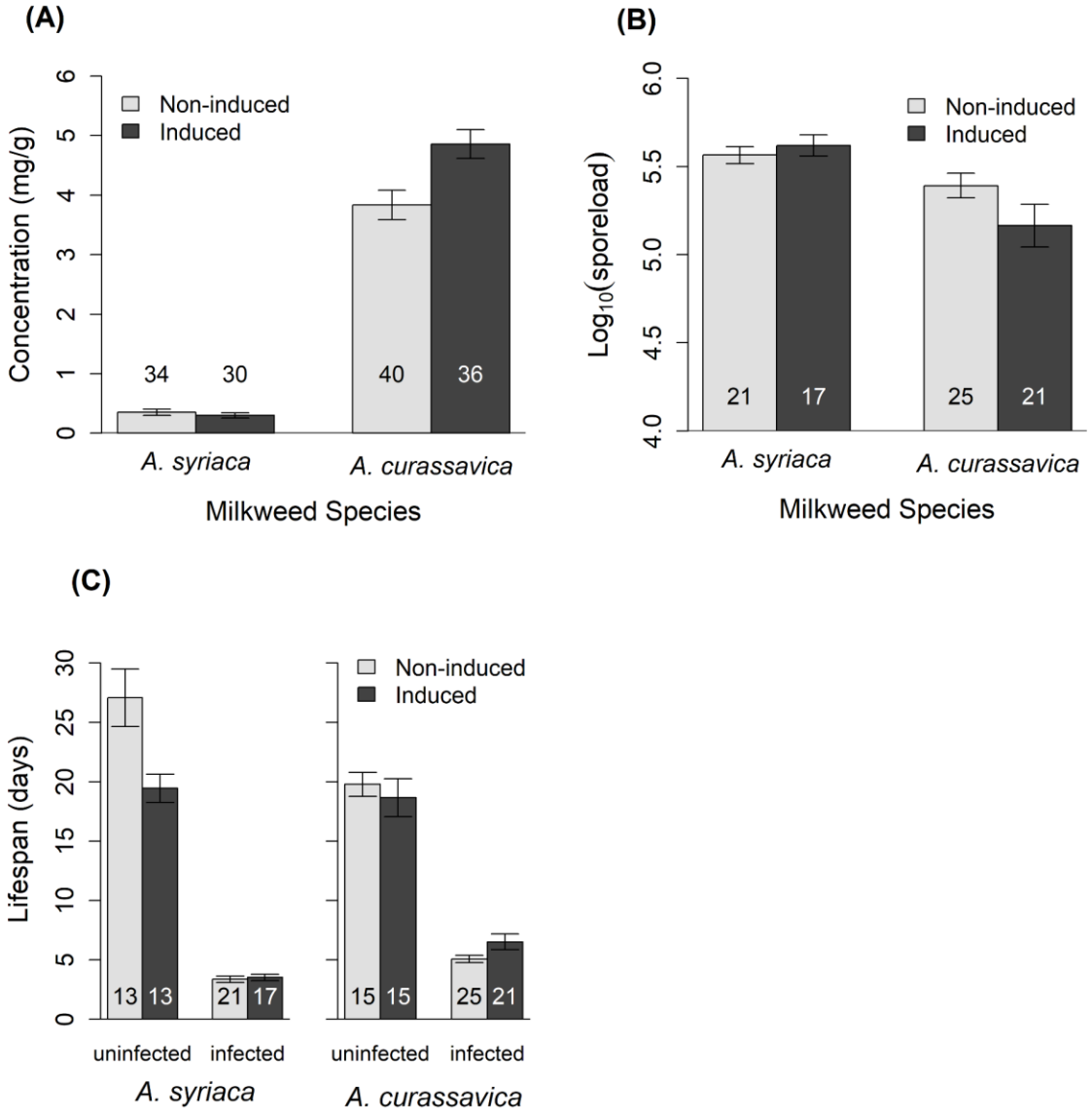


Figure 2.1. A) The effect of induction by monarch caterpillars on foliar cardenolide concentrations in two milkweed species; B) The effect of milkweed species and induction on parasite spore load in infected monarchs; C) The effect of milkweed species, induction, and parasite infection on the lifespan of monarchs. Data represent mean ±1 SEM. Sample sizes are reported on each bar.

2.4 Discussion

In this monarch-milkweed-parasite system, foliar cardenolides increase the resistance of monarchs to parasite infection (Sternberg et al., 2012). Aphid-induced declines in foliar cardenolides render monarchs more susceptible to parasites (de Roode, Rarick, et al., 2011). Here, we show that herbivory by monarch caterpillars induces a significant increase in foliar cardenolide concentrations in *A. curassavica* but not in *A. syriaca*. However, induction of cardenolides in *A. curassavica* did not translate into a statistically significant increase in monarch resistance to parasite infection. Previous work has shown that milkweed species vary strongly in cardenolide concentrations and in their effects on parasite resistance (*e.g.*, Sternberg et al., 2012); our current study suggests that these interspecific differences outweigh within-species variation due to monarch induction. Thus, while induction caused a 1.3-fold increase in cardenolide concentration in *A. curassavica*, the difference between the two species, which resulted in a significant difference in parasite resistance, was 11-fold (Fig. 2.1A). Other studies have also reported modest increases in cardenolide concentrations in response to monarch caterpillar grazing, and demonstrated that environmental factors such as light can alter the expression and magnitude of milkweed induced defenses (Agrawal et al., 2014; Agrawal, Kearney, et al., 2012). In addition, the age of plants could be a factor influencing the strength of phytochemical induction: some studies that observed cardenolide induction in *A. syriaca* used one-month old plants (Agrawal et al., 2014; Mooney, Jones, & Agrawal, 2008), while we used three-month old plants. The relatively small magnitude of cardenolide induction by monarchs in older plants could be the reason for the small effect sizes on parasite resistance observed in this study. It is important to realize that one-month old plants cannot support the entire larval development of a monarch caterpillar; as such, our results using large

three-month old plants are highly relevant for our understanding of cardenolide induction on parasite resistance in monarchs.

In addition to inducing cardenolides, herbivory by monarch caterpillars may also alter other forms of milkweed defense, including latex exudation and cysteine protease production, and/or the nutritional properties of foliage such as the carbon to nitrogen ratio (Rasmann, Johnson, & Agrawal, 2009). Nutritional properties such as phosphorus concentration also affect monarch lifespan under parasite infection (Tao et al., 2015). We observed reductions in the lifespan of monarchs reared on induced *A. syriaca*, even though cardenolide concentrations were not induced. This result suggests that induction altered other milkweed qualities aside from cardenolides, and that these changes affected monarch fitness. This result confirms that an array of defense mechanisms and nutritional properties respond to herbivore damage (Karban & Baldwin, 1997), and may influence multi-trophic interactions (Hunter & Schultz, 1993).

Overall, our study illustrates that monarch-induced increases in cardenolides, a class of anti-parasitic chemicals, are not large enough to increase monarch resistance to their parasites. These results suggest that in this milkweed-monarch-parasite tri-trophic system, interspecific variation in cardenolide concentration is a more important driver of parasite defenses than is plasticity via inducible defense.

2.5 Acknowledgments

We thank M. Tsai, A. Ahmad, and Y. Li for help with the experiments, H. Strait for help with the chemical analyses, H. Salem, Y. Yang, and two anonymous reviewers for providing constructive comments on the manuscript. This work was supported by National Science Foundation grant IOS-

1557724 to J.C.d.R and M.D.H.

Chapter 3: Transcriptomics of monarch butterflies reveals strong differential gene expression in response to host plant toxicity, but weak response to parasite infection

Wen-Hao Tan, Tarik Acevedo, Erica V. Harris, Tiffanie Y. Alcaide, James R. Walters, Mark D. Hunter, Nicole M. Gerardo, Jacobus C. de Roode

Abstract

Herbivorous insects have evolved many mechanisms to overcome plant chemical defenses, including detoxification and sequestration. Insect herbivores also may use toxic plants to reduce parasite infection. Plant toxins could interfere directly with parasites or could enhance endogenous immunity. Alternatively, plant toxins could favor down-regulation of endogenous immunity by providing an alternative (exogenous) defense against parasitism. However, studies on genome-wide transcriptomic responses to plant defenses and the interplay between hostplant toxicity and parasite infection remain rare. Monarch butterflies (*Danaus plexippus*) are specialist herbivores that feed on a range of milkweeds (*Asclepias* spp.), which contain toxic cardenolides. Monarchs have adapted to cardenolides through multiple resistance mechanisms and can sequester cardenolides to defend against bird predators. In addition, high-cardenolide milkweeds confer medicinal effects to monarchs against a common specialist protozoan parasite (*Ophryocystis elektroscirrha*). We used this system to study the interplay between the effects of host plant toxicity and parasite infection on global gene expression. Our results demonstrate that monarch larvae differentially express several hundred genes when feeding on *A. curassavica* and *A. incarnata*, two

milkweed species that are similar in nutritional content but that differ substantially in cardenolide concentrations. These differentially expressed genes include genes within multiple families of canonical insect detoxification genes, suggesting that they may play a role in monarch toxin resistance and sequestration. Interestingly, we found little transcriptional response to infection. However, parasite growth was reduced in monarchs reared on *A. curassavica*, and in these monarchs, a small number of immune genes were down-regulated, consistent with the hypothesis that medicinal plants can reduce reliance on endogenous immunity.

3.1 Introduction

Plants and herbivorous insects often have been used for studying coevolutionary arms races within the framework of chemical ecology (Rosenthal & Berenbaum, 1991). Plants have evolved many forms of defense against herbivores, such as the production of toxic secondary chemicals, and herbivorous insects have evolved mechanisms to overcome such plant defenses (Schoonhoven et al., 2005). These mechanisms include contact avoidance, rapid excretion, sequestration, enzymatic detoxification, and target site mutation (Després et al., 2007). Because host plants species vary in their secondary chemicals, herbivorous insects often utilize different mechanisms when feeding on different plants. For instance, milkweed aphids (*Aphid nerii*) differentially express several canonical insect detoxification genes, including genes encoding for Cytochrome P450s (CYP450s), UDP glucuronosyltransferases (UGTs), ATP-binding cassette transporters (ABC transporters), and Glutathione S-transferases (GSTs), when feeding on milkweed species that differ in toxicity (Birnbaum, Rinker, Gerardo, & Abbot, 2017). *Heliconius melpomene* also differentially express UGTs and GSTs when feeding on *Passiflora* species that differ in cyanogen content (Yu,

Fang, Zhang, & Jiggins, 2016). Herbivorous insects that feed on widely differing plant families have the additional complication that they may encounter an expanded range of phytochemicals, favoring plastic responses. Indeed, previous work has shown that the Swedish comma butterfly (*Polygonia c-album*) differentially expresses digestion- and detoxification-related genes, as well as genes encoding membrane transporters and cuticular proteins, when feeding on different host plant families (Celorio-Mancera et al., 2013).

While the ability to avoid, resist, or excrete toxic chemicals has been selected in many taxa, many insects also have evolved the ability to sequester secondary chemicals into their own tissues, thereby protecting themselves against their own natural enemies (Opitz & Müller, 2009). For example, in Lepidoptera (reviewed in Nishida, 2002), some swallowtail butterflies sequester aristolochic acid from their host plants to deter vertebrate predators (Uésugi, 2010); buckeye butterflies (*Junonia coenia*) sequester iridoid glycosides (IGs), which deter invertebrate predators (Dyer & Bowers, 1996; Theodoratus & Bowers, 1999); and tiger moths (*Grammia incorrupta*) sequester pyrrolizidine alkaloids, which defend them against parasitoids (Singer et al., 2009). In addition to the direct effects of sequestered chemicals on anti-predator and anti–parasite defense, phytochemicals also can affect parasites indirectly by modulating the host immune system (Lampert, 2012). Depending on the particular chemicals and parasites, toxin sequestration may reduce, enhance, or have no effect on anti-parasite immunity. For instance, all three scenarios have been shown in herbivores that sequester IGs. *Junonia coenia* exhibits reduced immunity (measured by the melanization response) when feeding on *Plantago lanceolata*, a plant species with greater concentrations of IGs, than when feeding on *P. major*, a less toxic host plant (Smilanich et al., 2009). In contrast, in this same system, feeding on the more toxic plant enhances anti-viral defenses (Smilanich et al., 2017). *Melitaea cinxia* shows enhanced immunity when feeding on *Plantago*

lanceolata strains with higher IG concentration (Laurentz et al., 2012), but in *Grammia incorrupta*, a moth species that also feeds on IG-containing plants, IG concentration does not appear to affect immune responses (Smilanich, Vargas, Dyer, & Bowers, 2011).

As described above, phytochemicals pose both challenges and benefits for herbivorous insects, and the ecological interactions and evolutionary relationships between plants and herbivorous insects have been studied extensively. However, studies of genome-wide transcriptomic responses to plant defenses, which provide insight into the simultaneous effects of toxins on detoxification, sequestration, and immune systems, remain rare (Celorio-Mancera et al., 2013; Vogel, Musser, & Celorio-Mancera, 2014). Even for herbivorous insect species with genomic and transcriptomic information available, transcriptomic research has rarely focused on herbivore-plant interactions (Vogel et al., 2014).

Here, we provide a transcriptomics-based analysis of parasite-infected and uninfected monarch butterflies (*Danaus plexippus*) feeding on different host plant species. Monarch butterflies are a prominent example of sequestration and aposematism (Agrawal, Petschenka, et al., 2012). Monarchs are specialist herbivores on milkweeds (mostly *Asclepias* spp.), but these plants vary widely in their toxicity, measured predominantly as the concentration and composition of cardenolides (Agrawal, Petschenka, et al., 2012). Cardenolides are steroids that are toxic to most animals because they inhibit the essential enzyme Na^+/K^+-ATPase that is responsible for maintaining membrane potentials (Agrawal, Petschenka, et al., 2012). Monarchs and other herbivorous insects specializing on cardenolide-containing plants have convergently evolved amino acid substitutions on the target site of the toxins that decrease binding affinity (Dobler, Dalla, Wagschal, & Agrawal, 2012; Zhen, Aardema, Medina, Schumer, & Andolfatto, 2012). Target site insensitivity largely enhances monarch resistance to cardenolides, but they are not completely

resistant to cardenolides (Agrawal, Petschenka, et al., 2012; Petschenka, Offe, & Dobler, 2012). There are fitness costs, including reduced larval survival and adult lifespan, for monarchs feeding on milkweed species with high cardenolide concentration or toxicity (Agrawal, 2005; Malcolm, 1994; Tao, Hoang, Hunter, & de Roode, 2016; Zalucki, Brower, & Alonso-M, 2001; Zalucki, Brower, & Malcolm, 1990; Zalucki & Brower, 1992). Despite these costs, monarchs have evolved the ability to sequester cardenolides into their own tissues, which, coupled with bright warning coloration, deters bird predators (Brower et al., 1968). In addition to the anti-predator protection provided by milkweeds, high-cardenolide milkweeds also provide protection against the common specialist parasite *Ophryocystis elektroscirrha* (de Roode, Pedersen, Hunter, & Altizer, 2008; Sternberg et al., 2012). Monarchs become infected with this parasite during their larval stage when ingesting parasite spores (Mclaughlin & Myers, 1970), but feeding on milkweeds with greater concentrations of cardenolides results in lower parasite infection, growth and virulence (de Roode, Pedersen, et al., 2008; de Roode, Rarick, et al., 2011; Gowler et al., 2015; Lefèvre et al., 2010; Sternberg, de Roode, & Hunter, 2015; Sternberg et al., 2012; Tan, Tao, Hoang, Hunter, & de Roode, 2018; Tao et al., 2015; Tao, Hoang, et al., 2016). At present, however, it remains unclear how cardenolides, parasites, and the monarch's immune system interact. On the one hand, cardenolides might interfere directly with parasites. This could result in a down-regulation of immune responses, as these chemicals would fulfill the same role as anti-parasitic immunity. Alternatively, cardenolides could stimulate the monarch immune system and thus enhance immune responses against parasites. Therefore, monarchs provide an excellent model to study how detoxification, toxin sequestration, and immunity interact in a system with a known association between phytochemicals and disease resistance.

In this study, we assessed differential gene expression between monarch larvae feeding on the

low-cardenolide *A. incarnata* and the high-cardenolide *A. curassavica* when infected or uninfected with the specialist parasite *O. elektroscirrha*. Specifically, we performed RNA-Seq on two tissue types of parasite-infected and uninfected larvae fed with either plant species. In addition, we quantified parasite resistance of the same batch of larvae and measured foliar cardenolide concentration in the same batch of milkweeds. While we found a limited transcriptional response to parasite infection, our results reveal a large number of genes that are differentially expressed in monarchs reared on the two milkweed species, including the down-regulation of four immune genes when fed on the high-cardenolide *A. curassavica*.

3.2 Materials and Methods

3.2.1 Monarchs, milkweeds, and parasites

Monarch butterflies in this study were obtained from a lab-reared, outcrossed lineage generated from wild-caught migratory monarchs collected in St. Marks, Florida, USA. The parasite clone (C_1-E_{25}-P_3) was isolated from an infected, wild-caught monarch from the same population. We used two species of milkweed in this study: *A. incarnata* and *A. curassavica*.

These two species were chosen because they are similar in nutrient content, but differ substantially in their level of cardenolides (toxic, secondary compounds); concentrations in *A. curassavica* are generally at least 10-fold higher than are those in *A. incarnata*. As a consequence, the milkweeds have been shown repeatedly to differentially affect monarch resistance to parasitism, with *A. curassavica* reducing parasite infection, growth, and virulence relative to *A. incarnata* (de Roode, Pedersen, et al., 2008; de Roode, Rarick, et al., 2011; Lefèvre et al., 2010; Sternberg et al., 2015, 2012; Tao et al., 2015; Tao, Hoang, et al., 2016). Milkweed seeds were obtained from Prairie

Moon Nursery (Winona, MN, USA). All milkweeds in this study were grown in a greenhouse under natural light conditions with weekly fertilization.

3.2.2 Experimental design and sample collection

We used second instar larvae for transcriptome sequencing, because larvae most likely become infected with *O. elektroscirrha* during early instars under natural conditions, through either vertical or horizontal transmissions (Altizer, Oberhauser, & Geurts, 2004; de Roode et al., 2009). We could not use first instars due to size limitations. Also, second instar larvae sequester the highest amounts of cardenolides relative to their body mass (Jones, Peschenka, Flacht, & Agrawal, 2019). Upon hatching, we reared larvae individually in petri dishes and fed them either *A. incarnata* or *A. curassavica*. We inoculated second instar larvae by adding ten parasite spores to an 8-mm diameter leaf disk taken from the milkweed species upon which they had been feeding, following an established protocol (de Roode, Yates, et al., 2008). Uninfected controls received leaf disks without spores. After larvae consumed their entire leaf disk, they were provided leaves of the same milkweed species *ad libitum.* Eighteen to twenty-four hours after parasite inoculation, we placed larvae in RNAlater and stored them at 4°C. We dissected all larvae within four days of collection. We separated the entire digestive tract (hereafter, gut) and the remaining body (hereafter, body) and put the samples into separate tubes with RNAlater. We stored these samples at -80 °C. Sample sizes for each treatment group and tissue type were provided in supplemental information Table S3.1.

We reared another subset of parasite-infected and uninfected larvae to adulthood on each plant species to quantify parasite resistance (N = 9-17 per treatment group). After parasite inoculation, larvae were transferred to individual rearing cups (473 mL) and fed leaves from either *A. curassavica* or *A. incarnata*. After pupation, pupae were placed in a laboratory room maintained at

25 °C under 14/10h L/D cycle. After eclosion, adults were placed in 8.9 x 8.9 cm glassine envelopes without a food source at 12 °C under 14/10h L/D cycle. Parasite load was quantified using a vortexing protocol described in de Roode et al., (2008). Normality and variance homogeneity were checked with the Shapiro-Wilk normality test and Fligner-Killeen test. Parasite spore load data were analyzed using a two-sample t-test. All analyses were performed in R version 3.5.2 (R Core Team, 2018).

3.2.3 Chemical analyses

We collected two types of samples for chemical analyses: milkweed foliage and larval frass. We collected foliage samples to confirm the differences in total cardenolide concentration between the two hostplant species. In addition, we collected larval frass to compare the differences between cardenolide composition before and after larval digestion. Foliage samples of the two plant species (N = 11-12 individual plants per species) were collected on the same day that we performed parasite inoculations. One leaf from the fourth leaf pair on each plant was chosen. Six leaf disks (424 mm^2 total) were taken with a paper hole punch from one side of the leaf and placed immediately into a 1 mL collection tube with cold methanol. Another six identical leaf disks were taken from the opposite side of the same leaf to measure sample dry mass. Frass samples, each from an individual larva, were collected from another subset of second instar larvae that were reared from hatch on *A. curassavica* (N = 17). We only focused on *A. curassavica*, because *A. incarnata* foliage contains very few cardenolides. Frass samples for each individual were collected for 24 hours during the second instar. Frass was collected into a 1 mL collection tube with cold methanol the same day it was produced. Total cardenolide concentrations and cardenolide compositions were analyzed using reverse-phase ultra-performance liquid chromatography (UPLC; Waters Inc., Milford, MA, USA)

following established methods (Tao et al., 2015). The absorbance spectra were recorded from 200 to 300 nm with digitoxin used as an internal standard. Under reverse-phase UPLC, cardenolide retention time decreases as polarity increases. For the plant samples, we analyzed the difference in total cardenolide concentration between the two species. Normality and variance homogeneity were checked with the Shapiro-Wilk normality test and Fligner-Killeen test. Cardenolide data were analyzed using a Mann–Whitney U test due to violation of assumptions of normality and variance homogeneity. All analyses were performed in R version 3.5.2 (R Core Team, 2018). Due to the complication of standardizing cardenolide quantities between foliage and frass samples, we assessed the differences in cardenolide compositions by comparing the cardenolide peaks between the two sample types.

3.2.4 RNA extraction, library preparation, and sequencing

We extracted total RNA from either gut or body tissues using the RNeasy RNA mini extraction kit (Qiagen) following the manufacturer's protocol. The quality and quantity of RNA samples were assessed using a nanodrop and bioanalyzer. Total RNA was sent to BGI (Beijing Genomics Institute, Hong Kong) for library preparation and sequencing. We sequenced the two tissue types (gut and body separately) of infected and uninfected larvae fed with either *A. incarnata* or *A. curassavica,* with 3-4 biological replicates per treatment (see supplemental information Table S3.1). We performed 50 bp single-end sequencing with a sequencing depth of 20M reads per sample using the BGIseq-500 platform.

3.2.5 Transcriptome assembly

We checked the quality of RNA-seq reads using FastQC (Andrews, 2010) and compiled across

samples using MultiQC (Ewels, Magnusson, Lundin, & Käller, 2016). Sequence quality was consistently high across positions (see supplemental information Fig. S3.1), so we proceeded without trimming. RNA-seq reads for each sample were mapped to the monarch reference genome (Zhan, Merlin, Boore, & Reppert, 2011) using STAR ver 2.5.2b (Dobin et al., 2013) and checked for alignment statistics. There were two samples that had low quality; one of them had a very low quantity of reads and the other had a very low mapping rate. Given that these two samples were from different individuals, we removed four samples (i.e., both tissue types of the same individual) from our analyses. We obtained the number of reads mapped to each gene from STAR and compiled them across samples as a count matrix.

3.2.6 Differential gene expression analysis

Differential gene expression analysis was performed using the R Bioconductor package edgeR version 3.24.3 (Robinson, McCarthy, & Smyth, 2009). We performed separate analyses on the two tissue types. We removed genes without any counts across samples from our analyses. We normalized the library sizes across samples using the trimmed mean of M-values (TMM) normalization. We performed differential gene expression analyses using negative binomial generalized linear models (GLMs). We created design matrices for GLM with infection treatment and plant species as factors, estimated dispersion parameters, and fitted the models. We addressed specific questions of interest by setting coefficient contrasts to compare between different treatment groups. First, we compared gene expression between all infected and all uninfected larvae to examine the overall impacts of parasite infection. We then compared gene expression between infected and uninfected larvae reared on the two milkweeds species separately to examine plant-specific effects. Next, we compared gene expression between larvae fed with *A. incarnata* and *A.*

curassavica; given that we found almost no differences between infected and uninfected groups, we combined them for this comparison. The Benjamini-Hochberg method (Benjamini & Hochberg, 1995) was used to account for multiple hypothesis testing and to calculate the adjusted p-values. We visualized the results through heatmaps with hierarchical clustering, MA plots, and volcano plots generated using the R package edgeR version 3.24.3 (Robinson et al., 2009) and gplots version 3.0.1 (Warnes et al., 2016). All analyses were performed in R version 3.5.2 (R Core Team, 2018).

3.2.7 Examine specific gene sets of interest

Given that we were specifically interested in genes that function in immunity and detoxification, we examined if canonical immune genes and detoxification genes were differentially expressed among treatment groups. We obtained a full set of annotated monarch immune genes published by the *Heliconius* Genome Consortium (2012), which included a set of annotated (*Heliconius*) immune genes and their orthologs in several species, including monarchs. The monarch orthologs listed in this published dataset were based on a previous version of monarch genome annotation (OGS1.0), so we updated this full set of immune genes to the latest version of gene annotation (OGS2.0) using information provided in Monarch Base (Zhan & Reppert, 2013). This updated monarch immune gene set contains 114 genes belonging to the functional classes of recognition, signaling, modulation, and effector (see supplemental information Table S3.2). For detoxification genes, similar to a previous study on another milkweed-feeding insect (Birnbaum et al., 2017), we focused on four canonical gene families: Cytochrome P450s (CYP450s), UDP glucuronosyltransferases (UGTs), ATP-binding cassette transporters (ABC transporters), and Glutathione S-transferases (GSTs). We obtained those annotated detoxification genes from

Monarch Base (Zhan & Reppert, 2013). We examined each set of our significantly differentially expressed genes to obtain the number of immune and detoxification genes within them. For all the significantly differentially expressed detoxification genes, we performed BLAST searches against two other Lepidopteran species (*Bombyx mori* and *Heliconius melpomene*) via the EnsemblMetazoa database (https://metazoa.ensembl.org/) to verify that their top hit paralogs also have the same putative detoxification function.

3.2.8 Gene ontology enrichment analysis

Functional annotations and Gene Ontology (GO) term assignments for all protein coding genes in the genome were generated using PANNZER2 (Törönen, Medlar, & Holm, 2018), with protein sequences obtained from Monarch Base, using default parameters. We created a custom annotation package for our organism using AnnotationForge (Carlson & Pages, 2018). We performed GO-term enrichment analyses on differentially expressed genes using ClusterProfiler (Yu, Wang, Han, & He, 2012) with default p-value and q-value cutoff thresholds. The lists of significantly differentially expressed genes were identified previously (see section 3.2.6). The "gene universe" included all genes that were expressed in our RNA-Seq dataset. The Benjamini-Hochberg method (Benjamini & Hochberg, 1995) was used to account for multiple hypothesis testing and to calculate the adjusted p-values. We included all three ontology groups in our analyses: biological process (BP), molecular function (MF), and cellular components (CC). We visualized the enrichment results by dotplots using ClusterProfiler (Yu et al., 2012)

3.3 Results

3.3.1 Plant chemistry and parasite resistance

We confirmed previous findings that the two milkweed species differ greatly in cardenolide concentration and differentially affect monarch resistance to parasitism. Total cardenolide concentration of *A. curassavica* foliage was 95-fold higher than that of *A. incarnata* foliage (Fig. 3.1A; $W = 0$, $P < 0.0001$), and butterflies reared on *A. curassavica* experienced significantly lower parasite spore load than those fed with *A. incarnata* (Fig. 3.1B; $t = 3.39$, $df = 19$, $P = 0.003$). None of the uninoculated monarchs became infected (N = 9 for *A. incarnata* and N = 17 for *A. curassavica*). When comparing the cardenolide composition of *A. curassavica* foliage and the frass from larvae feeding on *A. curassavica*, we found that they differed greatly in composition (Fig. 3.2). Specifically, out of a total of 22 unique cardenolides (i.e., individual bars in Fig. 3.2), only four occurred in both foliage and frass; eight cardenolides were exclusively found in foliage, and nine were exclusively found in frass. Additionally, there were more polar cardenolides in frass than in foliage, as indicated by lower retention times relative to a digitoxin internal standard (Fig. 3.2).

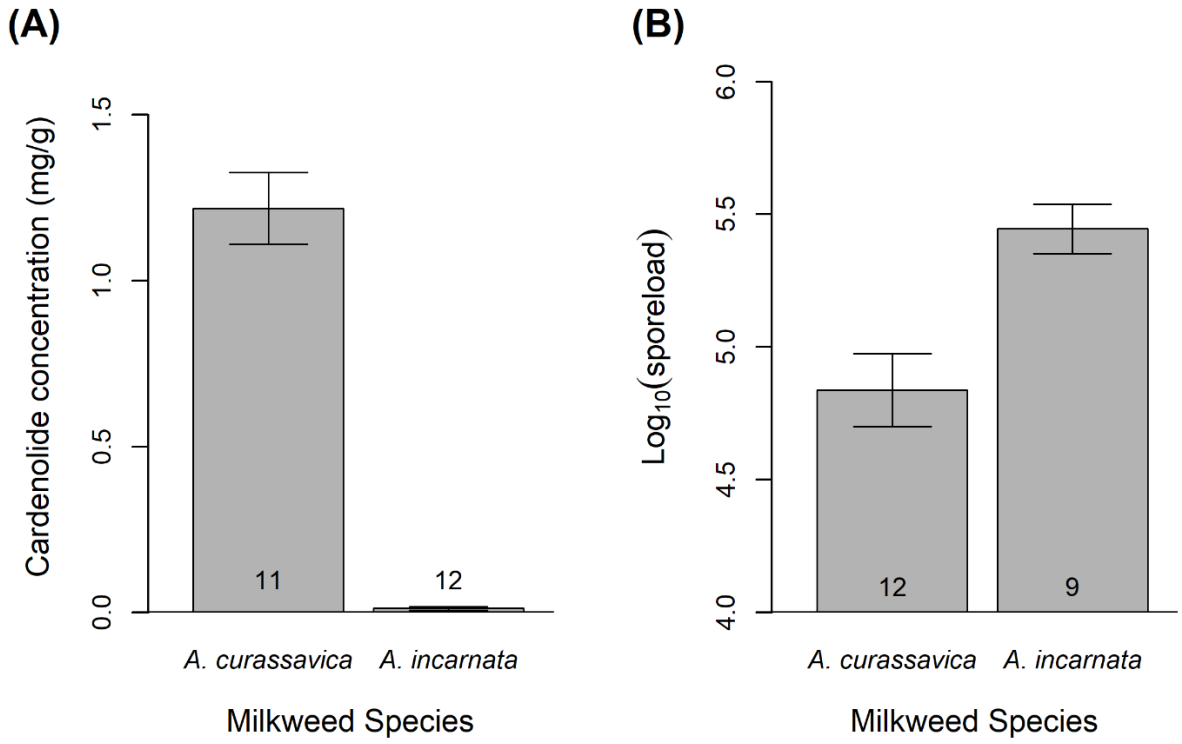


Figure 3.1. Differences in foliar cardenolide concentration and monarch parasite resistance between the two milkweed species, *A. curassavica* and *A. incarnata*. (A) Total cardenolide concentraion of foliage. (B) The effect of milkweed species on parasite spore load in infected monarchs. Data represent mean ±1 SEM. Sample sizes are reported on each bar.

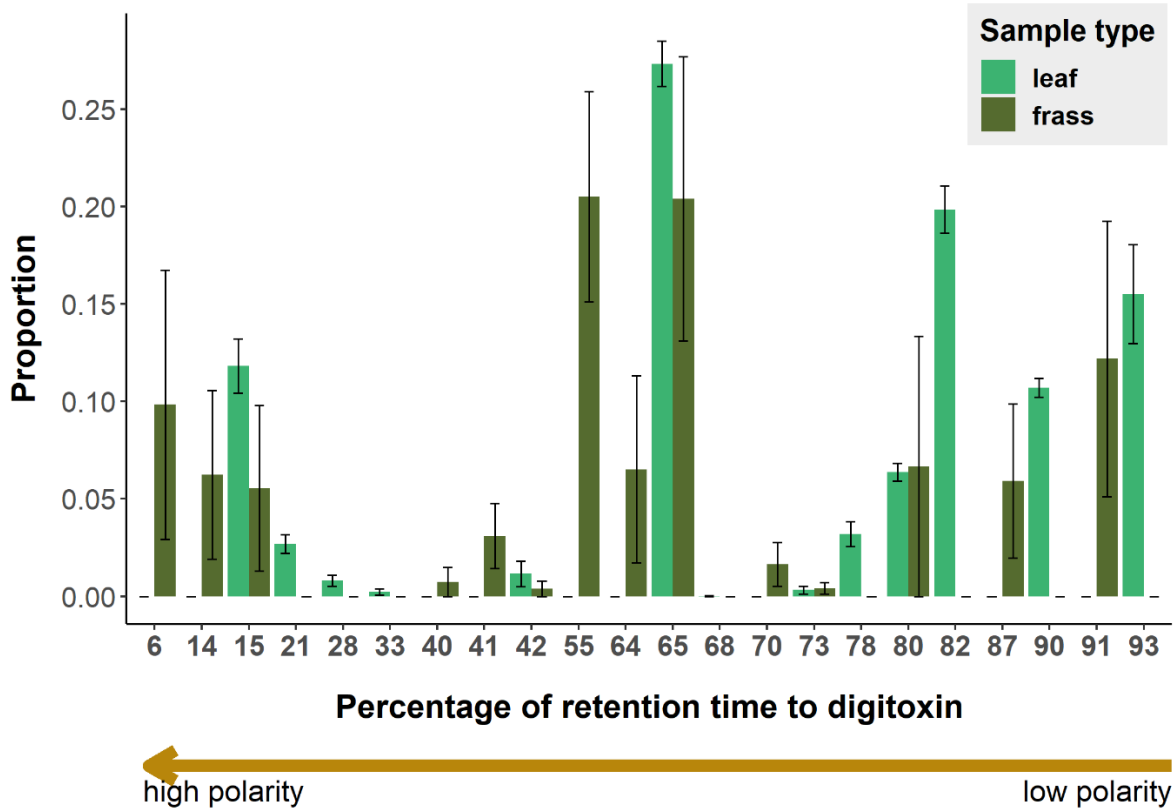


Figure 3.2. Cardenolide composition of *A. curassavica* foliage and frass produced by larvae fed with *A. curassavica*. The X-axis represents the percentage of retention time relative to a digitoxin internal standard in UPLC. Bars represent individual cardenolides. The Y-axis represents the proportion of the individual cardenolide within each sample. Data represent the mean ±1 SEM. Sample sizes: N = 11 for foliage samples (each sample was collected from a different individual plant) and N = 17 for frass samples (each sample was collected from a different individual larva). We only focused on *A. curassavica* because *A. incarnata* foliage contains very few cardenolides.

3.3.2 Differential gene expression analysis in relation to parasite infection

We first compared gene expression between all infected and all uninfected larvae to examine the overall effects of parasite infection on gene expression. Surprisingly, in both gut and body tissues, we found that no genes were differentially expressed (Fig. 3.3-3.4, Table 3.1). Next, we compared gene expression between infected and uninfected larvae reared on the two milkweed species separately to examine plant-specific effects. Again, we found almost no response to parasite infection (Table 3.1). For the larvae fed with *A. incarnata*, only one gene was significantly up-

regulated in the gut in the infected group: a cytochrome P450 gene (DPOGS205609). For the larvae fed with *A. curassavica*, only two genes were significantly down-regulated in the body in the infected group: an acid digestive lipase (DPOGS211626) and a carboxypeptidase (DPOGS211663). Overall, we found extremely few differentially expressed genes between infected and uninfected larvae regardless of tissue type or host plant, and none of those that were significantly differentially expressed were canonical immune genes.

Table 3.1. Summary of differentially expressed genes. The first two columns denote specific comparisons and the subset of samples used. The last three columns indicate the number of significantly up-regulated and down-regulated genes upon infection, or between those fed with different milkweed species, in either gut tissue or body. First, we compared infected and uninfected larvae in all samples to assess overall transcriptional patterns of parasite infection (*i.e.*, the first row). We then compared infected and uninfected larvae reared on the two milkweed species separately to examine plant-specific effects (*i.e*., the second and third rows). Next, we compared larvae fed with *A. incarnata* and *A. curassavica*. Given that we found almost no differences between infected and uninfected groups, we combined them for this comparison (*i.e.*, the fourth row).

Factor	Subset	Direction	Gut	Body
Infection	All	up-regulated in infected	0	0
		down-regulated in infected	0	0
Infection	*A. incarnata*	up-regulated in infected	1	0
		down-regulated in infected	0	0
Infection	*A. curassavica*	up-regulated in infected	0	2
		down-regulated in infected	0	0
Plant	All	increased in *A. curassavica*	271	122
		Increased in *A. incarnata*	637	306

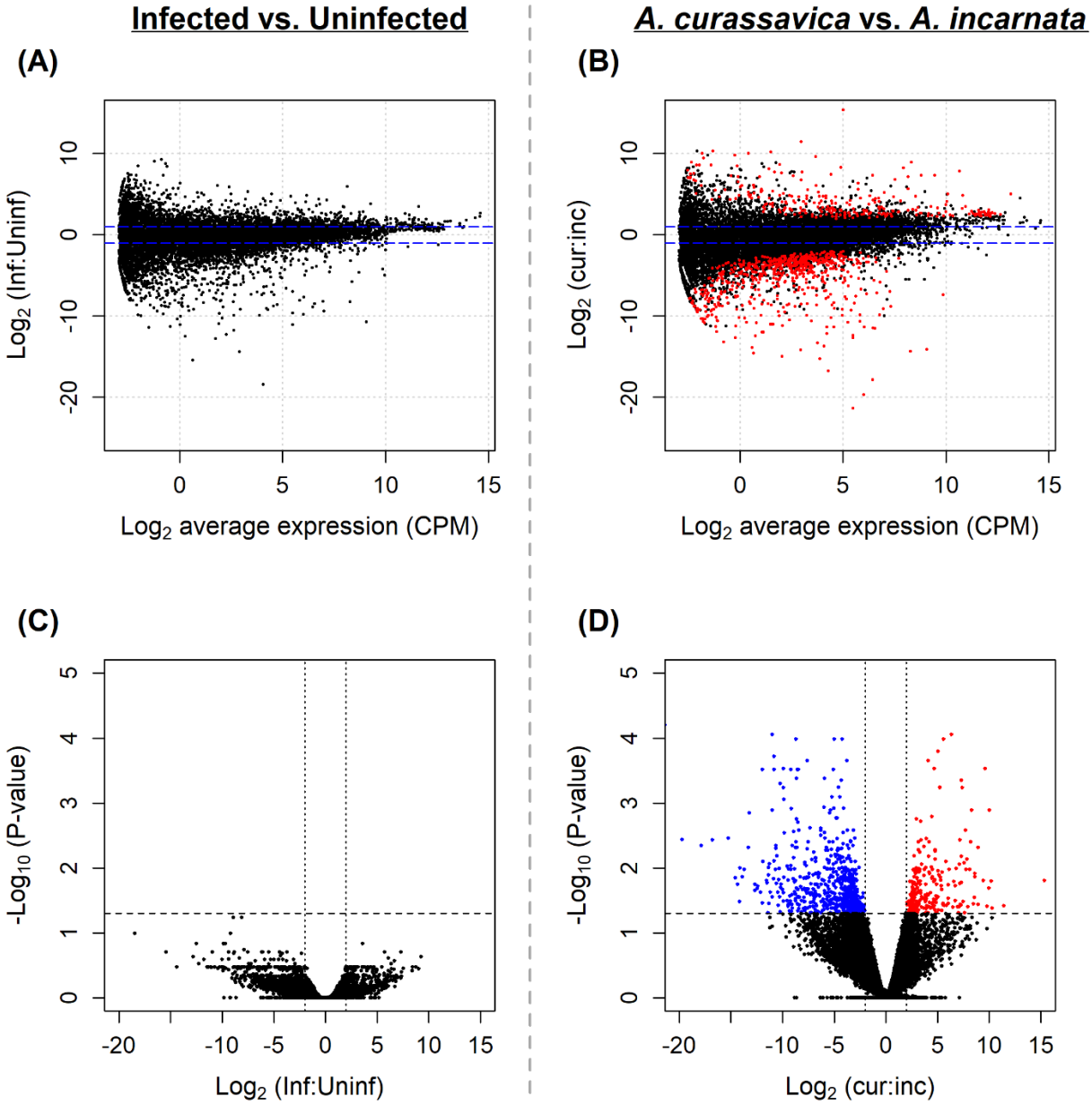


Figure 3.3. Patterns of differential gene expression in gut tissue. (A) and (C): expression differences between infected and uninfected larvae. A positive fold change indicates up-regulation in infected larvae. (B) and (D): expression differences between larvae fed with *A. curassavica* and *A. incarnata*. A positive fold change indicates up-regulation in larvae fed with *A. curassavica*. (A) and (B): MA plots. Dotted horizontal lines indicate ± 1-fold change. (C) and (D): volcano plots. Dotted horizontal lines indicate p-value thresholds. Dotted vertical lines indicate ± 2-fold change. Blue dots represent significantly down-regulated genes; red dots represent significantly up-regulated genes.

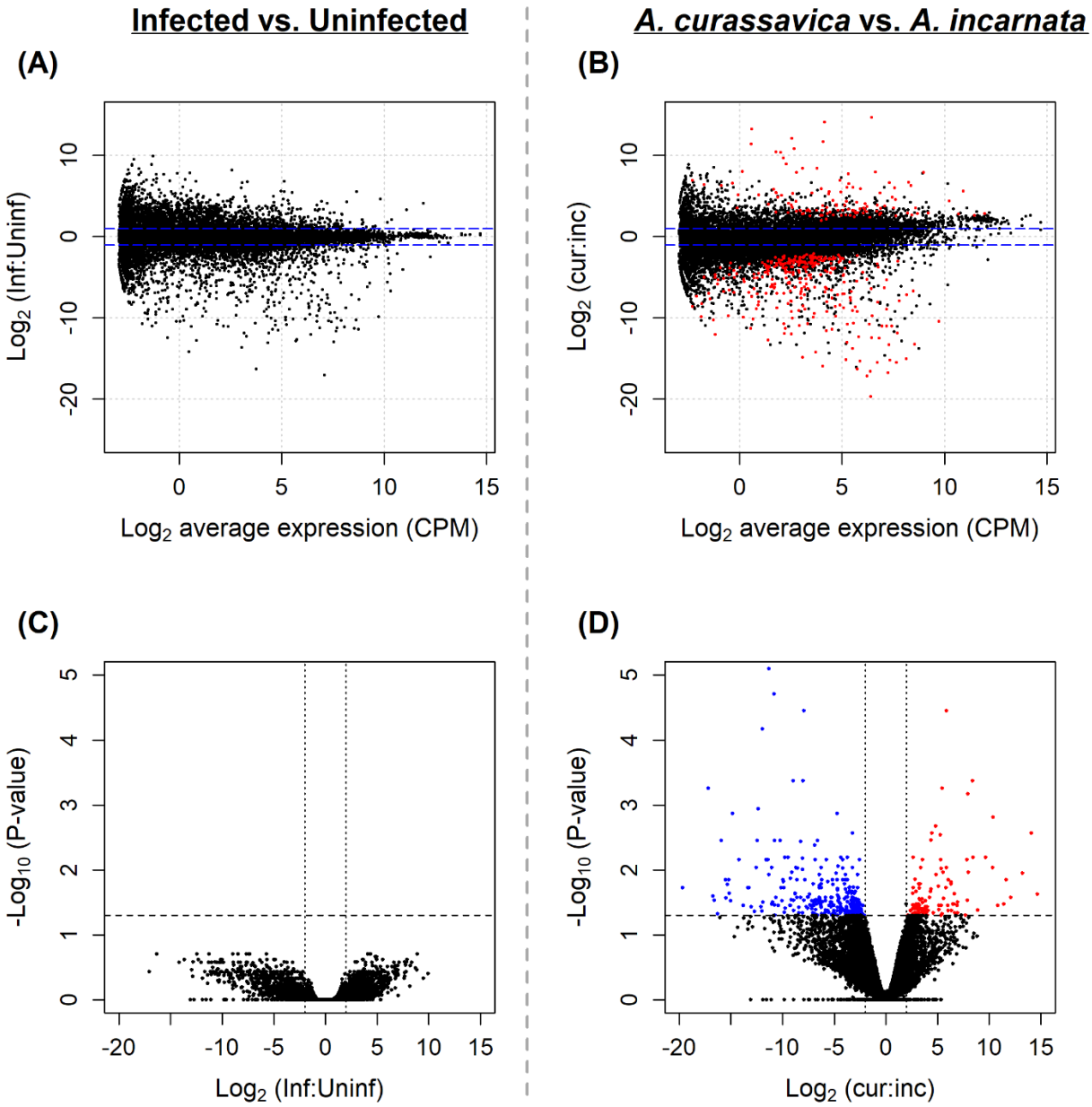


Figure 3.4. Patterns of differential gene expression in body tissue. (A) and (C): expression differences between infected and uninfected larvae. A positive fold change indicates up-regulation in infected larvae. (B) and (D): expression differences between larvae fed with *A. curassavica* and *A. incarnata*. A positive fold change indicates up-regulation in larvae fed with *A. curassavica*. (A) and (B): MA plots. Dotted horizontal lines indicate ± 1-fold change. (C) and (D): volcano plots. Dotted horizontal lines indicate p-value thresholds. Dotted vertical lines indicate ± 2-fold change. Blue dots represent significantly down-regulated genes; red dots represent significantly up-regulated genes.

3.3.3 Differential gene expression analysis in relation to milkweed diet

We compared gene expression between larvae reared on *A. curassavica* and *A. incarnata.* Given that we found almost no differences in expression between infected and uninfected larvae, we combined them in this comparison between plant species. We found that 908 genes were differentially expressed in the gut and 428 genes were differentially expressed in the body (Fig. 3.3-3.4, Table 3.1). Given that the gut is the place where initial digestion of plant matter happens, we expected the transcriptional patterns to be more distinct between plant diets in gut than in body samples. Indeed, heatmap and hierarchical clustering suggest that individuals are more clustered by plant diet in gut samples than in body samples (Fig. 3.5). The top 15 up-regulated and top 15 down-regulated genes for the gut and body are listed in Table 3.2 and Table 3.3, respectively. In gut tissues, notably, one of the top 15 up-regulated genes when fed with *A. curassavica* is a glutathione S-transferase (DPOGS210488), and another one is a carboxyl esterase (DPOGS204275), both of which are canonical insect detoxification genes. Other genes belong to a variety of biological functions, such as digestive processes and membrane-related proteins. Differential expression of digestive and membrane-related genes has also been demonstrated in other insects when feeding on different plant species (Celorio-Mancera et al., 2013).

In the body samples, three canonical detoxification genes were up-regulated when fed with *A. incarnata,* including one UDP-glycosyltransferase (DPOGS209528) and two cytochrome P450s (DPOGS207643 and DPOGS213243). In addition, the top 15 also includes a cytochrome b5 (DPOGS210599), which is a redox partner to cytochrome P450 in the P450 system (Després et al., 2007). Five of the top 15 up-regulated genes when fed with *A. curassavica* encode cuticular proteins. Interestingly, cuticle proteins have also been found to be differentially expressed in other insects when feeding on different host plants (Birnbaum et al., 2017; Celorio-Mancera et al., 2013).

Many of the remaining top differentially expressed genes (43.3% in gut and 30.0% in body) have unknown functions.

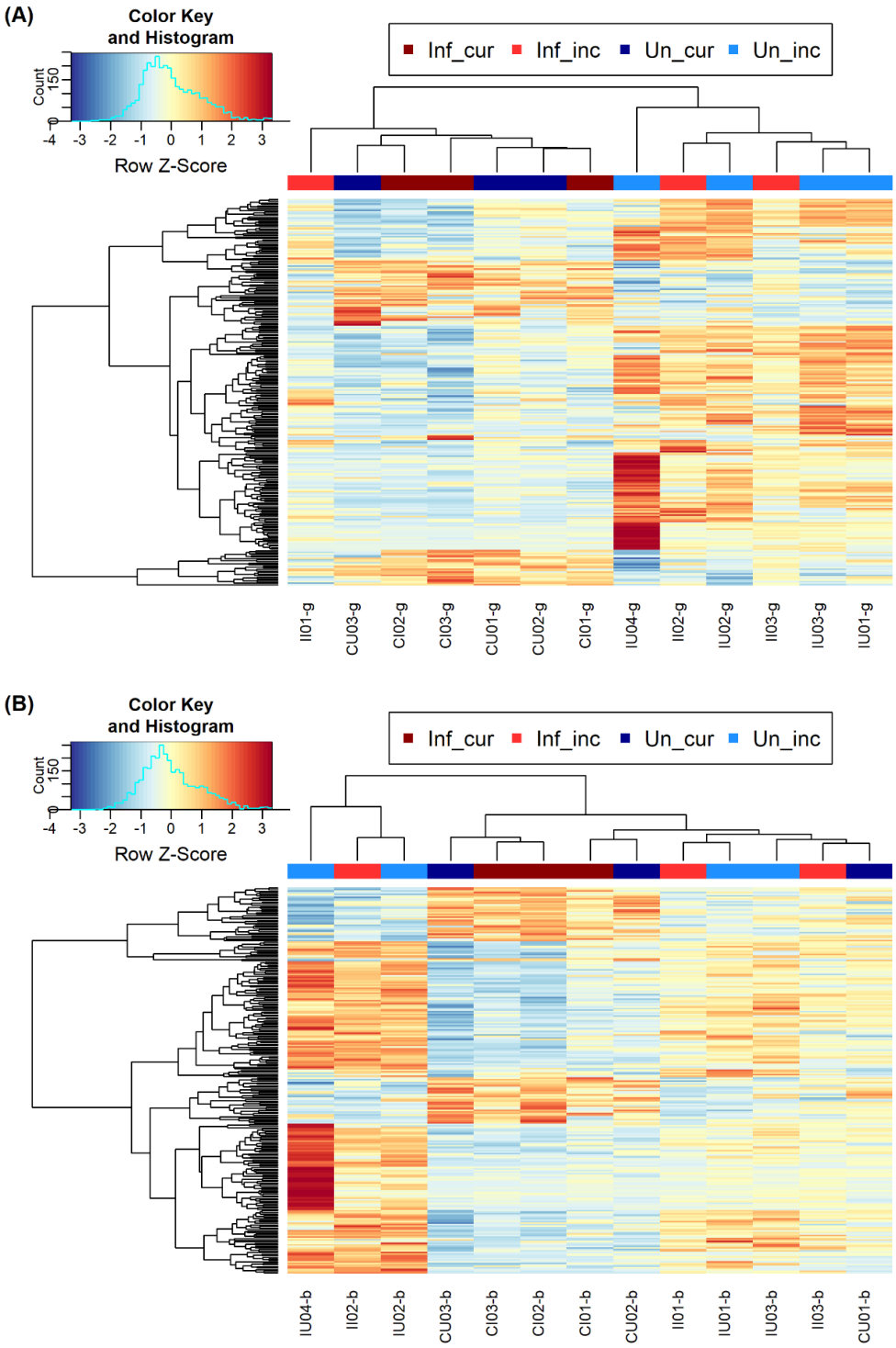


Figure 3.5. Heatmap and hierarchical clustering of the top 250 differentially expressed genes between larvae fed with *A. curassavica* and *A. incarnata*. (A) The result of gut samples. Hierarchical clustering shows that samples are clustered mostly based on the plant species larvae were fed with. (B) The result of body samples. The clustering patterns are less clear. "Inf_cur" represents infected larvae fed with *A. curassavica*; "Inf_inc" represents infected larvae fed with *A. incarnata*; "Un_cur" represents uninfected larvae fed with *A. curassavica*; "Un_inc" represents uninfected larvae fed with *A. incarnata*.

Table 3.2. List of top 15 differentially expressed genes in gut tissue between larvae fed with *A. curassavica* and *A. incarnata*. The list includes the top 15 genes significantly up-regulated when fed with *A. curassavica* and the top 15 genes significantly up-regulated when fed with *A. incarnata*.

Gene ID	log_2FC	logCPM	FDR	Protein
***Top 15 up-regulated genes in* A. curassavica**				
DPOGS201344	6.372	5.747	8.896E-05	Uncharacterized
DPOGS202254	5.589	5.739	1.040E-04	Threonine dehydratase catabolic-like isoform 2
DPOGS215709	5.049	13.155	1.596E-04	Uncharacterized
DPOGS212746	4.112	10.044	2.210E-04	Uncharacterized
DPOGS213427	4.699	4.654	2.947E-04	Phosphatidyltransferase
DPOGS204785	9.623	3.669	2.947E-04	Caboxypeptidase 4
DPOGS209145	7.309	6.446	4.455E-04	Uncharacterized
DPOGS204275	5.239	3.825	5.752E-04	Carboxyl/choline esterase
DPOGS213104	7.410	4.420	5.799E-04	Zinc finger protein
DPOGS204877	5.220	7.017	5.799E-04	Uncharacterized
DPOGS210488	10.030	-1.820	1.296E-03	Glutathione S-transferase epsilon 4
DPOGS205617	8.315	4.894	1.296E-03	Gucocerebrosidase
DPOGS200701	4.470	3.245	1.614E-03	Spliceosomal protein
DPOGS214834	2.985	6.014	1.746E-03	Juvenile hormone epoxide hydrolase
DPOGS206961	3.390	6.869	1.906E-03	Fructose 1,6-bisphosphate aldolase
***Top 15 up-regulated genes in* A. incarnata**				
DPOGS213127	-14.990	2.053	2.820E-06	Nuclear receptor GRF
DPOGS209249	-21.366	5.499	6.322E-05	Uncharacterized
DPOGS205455	-11.005	1.492	8.896E-05	Uncharacterized
DPOGS215049	-8.676	3.715	1.040E-04	Peroxidasin-like protein
DPOGS214337	-4.961	2.053	1.040E-04	Dystrophin
DPOGS206024	-4.189	4.407	1.040E-04	Uncharacterized
DPOGS205589	-10.789	5.246	1.909E-04	Hormone receptor 3C
DPOGS215508	-3.738	3.000	2.210E-04	Uncharacterized
DPOGS210943	-7.584	5.638	2.210E-04	Uncharacterized
DPOGS211620	-9.907	4.977	2.947E-04	Uncharacterized
DPOGS202595	-9.197	4.968	3.075E-04	Serpin-27
DPOGS209028	-8.462	1.370	3.075E-04	Uncharacterized
DPOGS207056	-10.801	0.320	3.075E-04	Uncharacterized
DPOGS200549	-5.086	1.072	3.075E-04	Aminopeptidase N-like protein
DPOGS200623	-8.542	2.970	3.075E-04	Molting fluid carboxypeptidase

Table 3.3. List of top 15 differentially expressed genes in body tissues between larvae fed with *A. curassavica* and *A. incarnata*. The list includes the top 15 genes significantly up-regulated when fed with *A. curassavica* and the top 15 genes significantly up-regulated when fed with *A. incarnata*.

Gene ID	log_2FC	logCPM	FDR	Protein
***Top 15 up-regulated genes in* A. curassavica**				
DPOGS202254	5.862	5.916	3.531E-05	Threonine dehydratase catabolic-like isoform 2
DPOGS207974	8.391	3.079	4.263E-04	Cuticle protein
DPOGS210599	5.474	4.955	5.561E-04	Cytochrome b5
DPOGS207878	7.965	6.632	6.747E-04	Antennal binding protein
DPOGS209820	10.405	1.778	1.544E-03	Allantoicase
DPOGS204877	4.834	7.461	2.096E-03	Neuropeptide-like precursor
DPOGS209878	14.095	4.153	2.685E-03	Cuticle protein
DPOGS201344	4.463	4.569	2.685E-03	Uncharacterized
DPOGS213427	5.256	4.346	2.893E-03	Phosphatidyltransferase
DPOGS212746	4.380	10.241	3.452E-03	Uncharacterized
DPOGS204901	8.429	3.785	6.396E-03	Cuticle protein
DPOGS202353	2.649	4.584	6.396E-03	Serine protease inhibitor 32
DPOGS200671	9.672	2.135	6.396E-03	Cuticle protein
DPOGS204876	5.325	2.603	6.911E-03	Uncharacterized
DPOGS204902	7.870	2.782	6.911E-03	Cuticle protein
***Top 15 up-regulated genes in* A. incarnata**				
DPOGS213127	-11.298	2.225	8.082E-06	Nuclear receptor GRF
DPOGS205589	-10.791	5.267	1.967E-05	Hormone receptor 3C
DPOGS216089	-7.901	2.515	3.531E-05	Uncharacterized
DPOGS209528	-11.924	2.200	6.803E-05	UDP-glycosyltransferase
DPOGS207933	-7.987	2.536	4.245E-04	Uncharacterized
DPOGS201723	-8.964	3.188	4.245E-04	Peritrophic matrix protein
DPOGS209249	-17.175	6.228	5.561E-04	Uncharacterized
DPOGS211620	-12.359	5.805	1.142E-03	Uncharacterized
DPOGS204937	-4.721	3.391	1.358E-03	Polypeptide N-acetylgalactosaminyltransferase
DPOGS212114	-14.837	3.068	1.358E-03	Laccase-like multicopper oxidase 2
DPOGS212041	-3.204	2.933	2.685E-03	Fibroblast growth factor receptor
DPOGS207643	-15.926	4.052	3.542E-03	Cytochrome P450 6AB4
DPOGS205455	-10.749	3.336	3.542E-03	Uncharacterized
DPOGS213243	-6.609	4.497	3.542E-03	Cytochrome P450
DPOGS201539	-12.438	6.447	3.542E-03	Uncharacterized

3.3.4 Examination of specific gene sets

Given existing evidence from other herbivore systems mentioned previously (Smilanich et al., 2009) and our hypothesis that host plants affect immune gene expression, we examined whether any of the known canonical insect immune genes were differentially expressed when feeding on different milkweed species. Among the full set of differentially expressed genes between larvae fed *A. curassavica* and *A. incarnata*, we found only four immune genes were significantly differentially expressed in gut tissue and only one immune gene was differentially expressed in whole-body tissue (Table 3.4). For the four differentially expressed immune genes associated with gut samples, two of them are CLIP serine proteases, one is a frep-like receptor, and the other one is a Toll-like receptor. The one differentially expressed gene associated with body samples is a CLIP serine protease that was also differentially expressed in the gut. Interestingly, all four of them were down-regulated in caterpillars fed *A. curassavica*, the more toxic species on which parasite growth was reduced. Overall, we did not find any support that more toxic milkweeds (i.e., *A. curassavica*) enhance the immunity of monarch larvae. Instead, we found weak support that feeding on more toxic milkweeds might cause down-regulation of a subset of immune genes.

Table 3.4. Canonical immune genes that were significantly differentially expressed in gut tissue between larvae fed with *A. curassavica* and *A. incarnata*. No canonical immune genes were significantly differentially expressed between infected and uninfected larvae.

Immune gene	Tissue	Direction	LogFC	LogCPM	FDR
CLIP serine protease (DPOGS215180)	gut	Increased in *A. incarnata*	-5.94	1.61	0.003
Frep-like receptor (DPOGS203317)	gut	Increased in *A. incarnata*	-4.85	1.79	0.007
CLIP serine protease (DPOGS213841)	gut	Increased in *A. incarnata*	-6.73	-0.47	0.012
Toll-like receptor (DPOGS211472)	gut	Increased in *A. incarnata*	-3.82	2.61	0.0140
CLIP serine protease (DPOGS215180)	body	Increased in *A. incarnata*	-5.76	2.45	0.04

Next, given that monarch larvae were fed with two milkweed species that differ greatly in the level of toxicity, we examined whether any of the known canonical insect detoxification genes were differentially expressed when feeding on the two milkweed species. We focused on gut tissues here because the gut is the place of primary contact with plant materials, where initial digestion and detoxification take place, and because we found stronger differential expression in gut than body tissues. We found that a large proportion of known detoxification genes were expressed (Table 3.5). Importantly, the proportion of detoxification genes within all significantly differentially expressed genes (2.42%) was significantly higher than the proportion of all annotated genes in the genome that are detoxification genes (1.35%) (χ^2 = 6.12, *df* = 1, *P* = 0.013), suggesting that they are overrepresented in the differentially expressed genes in monarchs reared on different milkweeds. The direction of differential expression was not universal, with some genes being up-regulated when larvae were reared on the toxic *A. curassavica* and others when reared on the less toxic *A. incarnata*. Specifically, 6 CYP450s, 2 UGTs, and 1 GST were up-regulated in monarchs fed *A. curassavica*, while 3 CYP450s, 1 UGTs, 8 ABC transporters, and 1 GST were up-regulated in

monarchs fed *A. incarnata* (Table 3.5 and Supplementary Table S3.3). Interestingly, all of the ABC transporters were only significantly up-regulated in monarchs fed with *A. incarnata*. Our results demonstrate that several canonical detoxification genes were differentially expressed when larvae fed on the two milkweeds species with different levels of toxicity, suggesting that these genes likely are involved in metabolizing secondary compounds.

Table 3.5. Canonical detoxification genes that were significantly differentially expressed in gut tissue between larvae fed with *A. curassavica* and *A. incarnata*. The second column, "Annotated", indicates the number of annotated genes in the genome for the given gene family. The third column, "Expressed", indicates the number of genes that were expressed in our RNA-seq dataset (defined as counts > 0 in at least two samples). The last two columns show the number of significantly differentially expressed genes.

Gene family	Annotated	Expressed	Increased in *A. curassavica*	Increased in *A. incarnata*
Cytochrome P450 (CYP)	75	72	6	3
UDP glucuronosyltransferases (UGT)	35	34	2	1
ATP-binding cassette transporters (ABC transporters)	61	60	0	8
Glutathione S-transferases (GSTs)	33	31	1	1

3.3.5 Gene ontology enrichment analysis

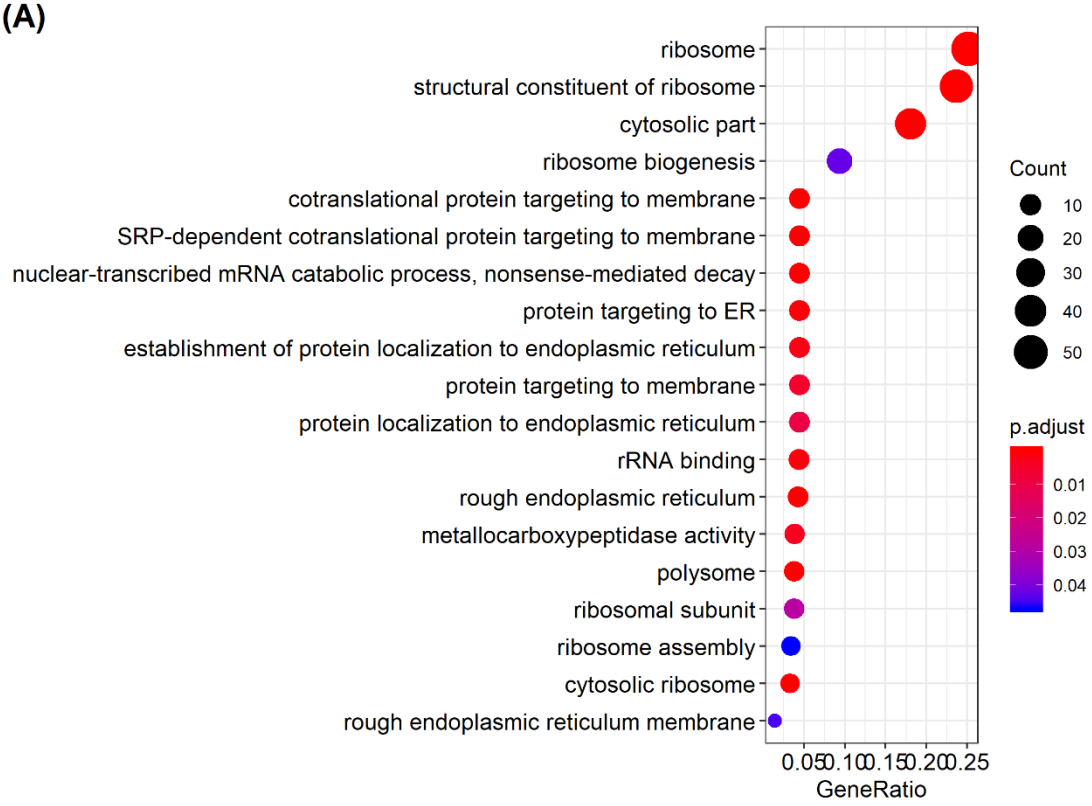
Given that there were almost no differentially expressed genes across infection treatments, we only performed GO enrichment analysis on differentially expressed genes between larvae fed with different plant species. We performed separate analyses for significantly up-regulated genes in larvae fed with *A. curassavica* and significantly up-regulated genes in larvae fed with *A. incarnata* in the two tissue types. Among up-regulated genes in larvae reared on *A. curassavica*, we found a total of 19 GO terms significantly enriched in the gut tissue and one GO term significantly enriched in the body. Among up-regulated genes in *A. incarnata*-reared larvae, we found a total of 112 GO

terms significantly enriched in the gut tissue and 6 GO terms significantly enriched in the body (Table 3.6). Significantly enriched GO terms for each group are shown in Fig. 3.6 & 3.7. Overall, we found many more significantly enriched GO terms in gut tissue than in body, and in larvae fed with *A. incarnata.* However, none of those GO terms have seemingly direct functional relevance to detoxification or immunity.

Table 3.6. Number of significantly functionally enriched GO terms in gut and body tissues between larvae fed with *A. curassavica* and *A. incarnata*. BP = biological process, MF = molecular function, CC = cellular component. Multiple testing was accounted for using the Benjamini-Hochberg method.

Tissue type	direction	BP	MF	CC	Total
Gut	Increased in *A. curassavica*	9	3	7	19
Gut	Increased in *A. incarnata*	102	0	10	112
Body	Increased in *A. curassavica*	0	1	0	1
Body	Increased in *A. incarnata*	4	2	0	6

(A)


ribosome
structural constituent of ribosome
cytosolic part
ribosome biogenesis
cotranslational protein targeting to membrane
SRP-dependent cotranslational protein targeting to membrane
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
protein targeting to ER
establishment of protein localization to endoplasmic reticulum
protein targeting to membrane
protein localization to endoplasmic reticulum
rRNA binding
rough endoplasmic reticulum
metallocarboxypeptidase activity
polysome
ribosomal subunit
ribosome assembly
cytosolic ribosome
rough endoplasmic reticulum membrane
0.05
0.10
0.15
0.20
0.25
GeneRatio
Count
10
20
30
40
50
p.adjust
0.01
0.02
0.03
0.04

(B)

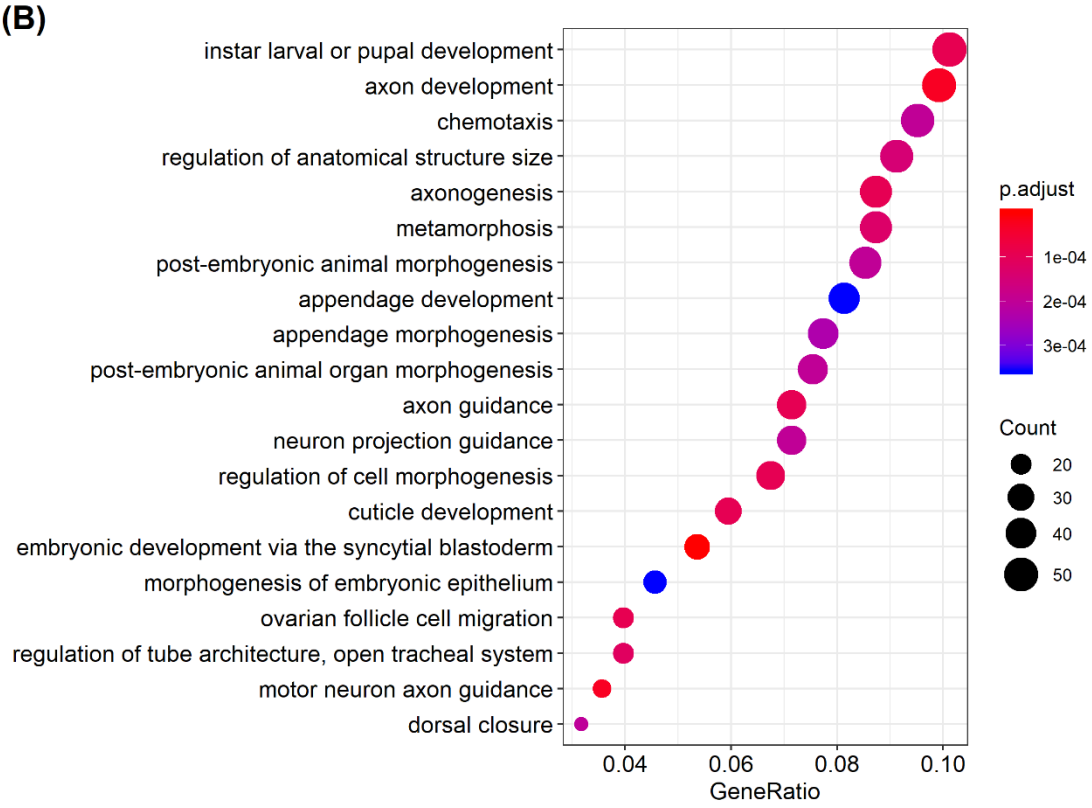

instar larval or pupal development
axon development
chemotaxis
regulation of anatomical structure size
axonogenesis
metamorphosis
post-embryonic animal morphogenesis
appendage development
appendage morphogenesis
post-embryonic animal organ morphogenesis
axon guidance
neuron projection guidance
regulation of cell morphogenesis
cuticle development
embryonic development via the syncytial blastoderm
morphogenesis of embryonic epithelium
ovarian follicle cell migration
regulation of tube architecture, open tracheal system
motor neuron axon guidance
dorsal closure
0.04
0.06
0.08
0.10
GeneRatio
p.adjust
1e-04
2e-04
3e-04
Count
20
30
40
50

Figure 3.6. Significantly functionally enriched GO terms in gut tissue between larvae fed with *A. curassavica* and *A. incarnata*. (A) 19 significant terms in up-regulated genes in *A. curassavica*. (B) 116 significant terms in up-regulated genes in *A. incarnata*. Only the top 20 were shown. The x-axis represents the proportion of genes that belong to a given functional category to the total number of differentially expressed genes. All three ontology terms (BP, MF, CC) were included. BP = biological process, MF = molecular function, CC = cellular component. P-values were corrected using the Benjamini-Hochberg method.

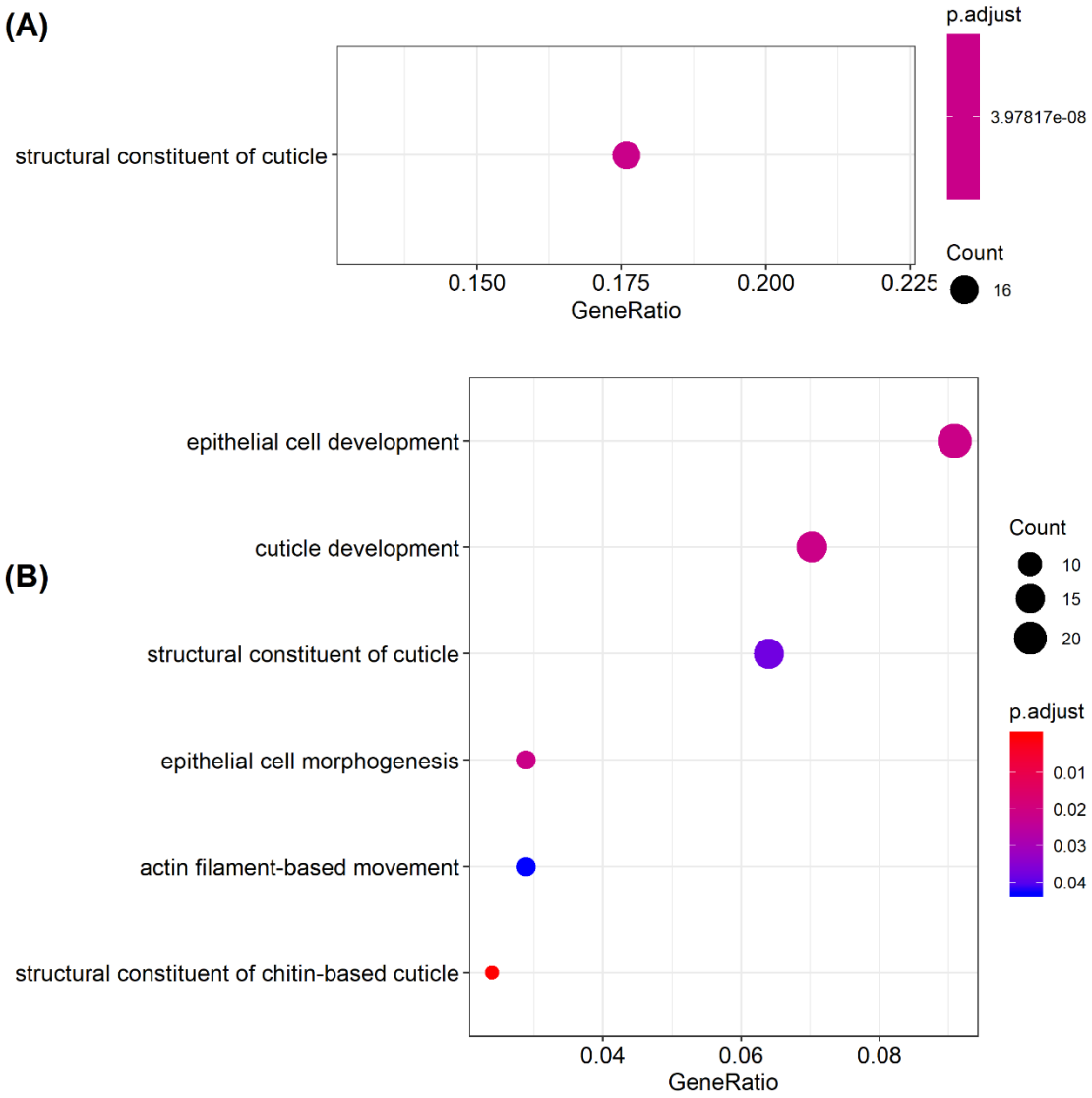


Figure 3.7. Significantly functionally enriched GO terms in body tissue between larvae fed with *A. curassavica* and *A. incarnata*. (A) One significant term in up-regulated genes in *A. curassavica*.

(B) Six significant terms in up-regulated genes in *A. incarnata*. The x-axis represents the proportion of genes that belong to a given functional category to the total number of differentially expressed genes. All three ontology terms (BP, MF, CC) were included. BP = biological process, MF = molecular function, CC = cellular component. P-values were corrected using the Benjamini-Hochberg method.

3.4 Discussion

This study examined differences in transcriptional profiles between monarch butterfly larvae feeding on two milkweed species and in response to infection by a specialist protozoan parasite. Our results demonstrate that hundreds of genes were differentially expressed in gut and body when feeding on two different milkweed species. Given that these two milkweed species differ greatly in their concentrations of secondary chemicals (cardenolides) (Fig. 3.1A), but little in nutrient composition (Tao, Ahmad, de Roode, & Hunter, 2016), these transcriptional differences are likely related to coping with different levels of toxicity in their diet. In fact, we found that several canonical insect detoxification genes were differentially expressed in monarchs reared on the two milkweed species. We discovered that many more genes were differentially expressed in gut than body tissue and that transcriptional profiles of gut samples formed more defined clusters, suggesting that transcriptional responses in relation to milkweed diet are stronger in the gut than in the rest of the body. We also found four canonical immune genes that were differentially expressed between individuals fed on different milkweed species. Interestingly, all four immune genes were down-regulated in monarchs reared on *A. curassavica*, the plant species that reduced parasite infection. In contrast with these transcriptional responses to milkweed diet, we found few transcriptional differences between infected and uninfected monarchs.

3.4.1 Detoxification of plant secondary chemicals

Many plants produce secondary metabolites as defense chemicals against herbivores. In response, herbivorous insects express genes that function in several protective mechanisms, including enzymatic detoxification, excretion, and sequestration (Després et al., 2007). Some previous studies have demonstrated that insects differentially express detoxification genes when feeding on plants with different levels of defense chemicals. For instance, *Drosophila mettleri*, a fruit fly species specialized on cacti with toxic alkaloids, differentially expresses several detoxification genes, including P450s, UGTs, GST, and carboxylesterases, when feeding on different food sources (Hoang, Matzkin, & Bono, 2015). *Tupiocoris notatu*s, a mirid species, down-regulates several GST, UGT, and P450s when feeding on defenseless (JA-silenced) *Nicotiana attenuata* (Crava, Brütting, & Baldwin, 2016). Similarly, our results demonstrate differences in transcriptional profiles of monarch larvae feeding on different milkweed species. Several of those differentially expressed genes belong to canonical detoxification genes, including P450s, UGTs, GSTs, and ABC transporters. Detoxification-related categories, however, were not significantly enriched in our enrichment analyses. While the majority of detoxification genes were expressed, only a relatively small proportion of them were differentially expressed between monarchs reared on the different plant species. Taken together, these results suggest that although a large number of detoxification genes are required for metabolizing a toxic plant diet, only a relatively small proportion of them are related to dealing with variable levels of toxicity. Although our significantly enriched expression categories are not related to detoxification, many of them have also been reported in other studies of herbivorous insects. For instance, categories related to membrane, cuticle, and ribosome are significantly enriched in *Polygonia c-album* when feeding on different plant species (Celorio-Mancera et al., 2013). Enrichment of cuticle-related genes when feeding on

different hostplants has also been reported in milkweed aphids (Birnbaum et al., 2017) and in some other herbivorous insects (Hoang et al., 2015; Matzkin, 2012). Thickening cuticular components have been suggested to reduce the penetration of insecticides, facilitating insecticide resistance (Foster et al., 2010). Alternatively, as certain insecticides are known to inhibit chitin synthesis (Leighton, Marks, & Leighton, 1981), it is possible that insects regulate the transcription of cuticle-related genes to deal with the interference of plant toxins on chitin metabolism and cuticular protein interactions (Celorio-Mancera et al., 2013).

CYP450 is one of the largest gene families in insects and catalyzes a wide range of reactions (Werck-Reichhart & Feyereisen, 2000). In many insects (e.g., black swallowtail (*Papilio polyxenes*) and parsnip webworm (*Depressaria pastinacella*)), the monooxygenase activity of P450s plays an important role in metabolizing plant toxins such as furanocoumarins (Mao, Rupasinghe, Zangerl, Schuler, & Berenbaum, 2006; Schuler, 1996; Wen, Pan, Berenbaum, & Schuler, 2003). Cardenolides are also substrates for CYP450 monooxygenases (Marty & Krieger, 1984), and it is assumed that milkweed-feeding insects metabolize cardenolides during the detoxification process (Agrawal, Petschenka, et al., 2012). Our results indicate that many CYP450 genes are expressed and some of them are differentially expressed when feeding on milkweeds with different levels of cardenolides, suggesting that they possibly play a role in detoxifying cardenolides. Furthermore, our chemical analyses comparing foliage and frass cardenolide composition identified specific cardenolides in frass that are not present in foliage, including several with high polarity. This result, consistent with a recent study (Jones et al., 2019), suggests that some of the cardenolides excreted via frass are likely modified forms, created through detoxification processes. We hypothesize that CYP450 genes may play a role in this modulation.

3.4.2 Specialization on cardenolide-containing plants and sequestration of cardenolides

Despite the fact that milkweed-feeding insects have been one of the most studied systems in chemical ecology and plant-insect interactions, to our knowledge, very few studies have characterized global transcriptional responses of specialist insects when feeding on milkweeds. Recently, Birnbaum *et. al.* (2017) compared transcriptional profiles using both RNA-seq and qPCR of milkweed aphids (*Aphid nerii*) fed on three different milkweed species, including the plant species used in our study. Similar to our study, they found differential expression of canonical insect detoxification genes, including genes belonging to CYP450s, UGTs, GSTs, and ABC transporters. In addition, their findings and our results both indicate that a greater number of genes are down-regulated rather than up-regulated when milkweed-specialized insects feed on more toxic plant species (Table 3.1)(Birnbaum et al., 2017). Although both studies on milkweed-feeding insects showed similar results, milkweed aphids do not have the target site mutations on Na^+/K^+-ATPase that confer resistance to cardenolides in monarchs (Zhen et al., 2012), suggesting that they rely on other mechanisms to cope with cardenolides. A previous study across three milkweed-feeding butterflies that differ in target site sensitivity indicated that resistance conferred by target site insensitivity has a stronger association with sequestering cardenolides than with digesting cardenolide-rich diets (Petschenka & Agrawal, 2015). Therefore, since the two species differ in target site sensitivity, but exhibit similar transcriptional responses to feeding on more toxic plants, the differentially expressed genes may be important in sequestration processes, as both species sequester cardenolides as a defense against predators (Rosenthal & Berenbaum, 1991).

Previous studies have demonstrated that monarch larvae can regulate the level of cardenolide sequestration, as indicated by the fact that cardenolide concentration in larval hemolymph and milkweed leaves do not show a linear relationship (Rosenthal & Berenbaum, 1991). Interestingly,

monarchs concentrate cardenolides when feeding on low-cardenolide plants and sequester less when feeding on plants with a very high concentration of cardenolides (Jones et al., 2019; Malcolm, 1991). Notably, our results show that all the differentially expressed ABC transporters were up-regulated in larvae fed *A. incarnata*, a milkweed species with very low cardenolide concentrations. Studies of other insect systems have shown that ABC transporters are involved in sequestration processes. For example, ABC transporters play a key role in salicin sequestration in poplar leaf beetles (*Chrysomela populi*) (Strauss, Peters, Boland, & Burse, 2013). Therefore, the up-regulation of ABC transporters when feeding on low-cardenolide milkweed might be related to an increased rate of cardenolide sequestration.

3.4.3 The effects of plant diet on immunity

Some studies have demonstrated that plant diets with high toxicity can reduce immune responses of herbivorous insects (Smilanich et al., 2009). Detoxification and sequestration of plant toxins can be energetically costly (Bowers, 1992), so a reduction in immune function could be caused by trade-offs with these processes (Moret & Schmid-Hempel, 2000). Plant toxins may have direct negative effects on immune cells (Smilanich et al., 2009). Alternatively, insect hosts may invest less in immunity when anti-parasite resistance is provided by host plants instead. In our study, although we did not find a strong overall effect of plant diet on the expression of canonical immune genes, we observed reduced expression of four immune genes in monarchs feeding on *A. curassavica*, the anti-parasitic plant species. This does not preclude the possibility that other monarch immune defenses not captured by gene expression differences may be influenced by host plant diet. Future studies should couple investigation of immune gene expression with studies of cellular immune responses and should strive to characterize the function of the many genes of

unknown function in monarchs, some of which could play a role in defense.

In the context of herbivore-parasite interactions, medicinal effects conferred by plant diet could be mediated by either direct or indirect effects of plant toxins on parasites. Specifically, medicinal compounds may directly interfere with parasites or may indirectly enhance disease resistance by stimulating immune responses. In the former scenario, investment in immune responses may be reduced because they are compensated for by the medicinal compounds. Indeed, recent studies have demonstrated that the use of medicinal compounds reduces immune investment in a variety of insect species. For example, honey bees (*Apis mellifera*) provided with resins, which have antimicrobial properties, exhibit reduced expression of two immune genes (Simone et al., 2009). Similarly, the presence of resins also reduces humoral immune responses in wood ants (*Formica paralugubris*) (Castella et al., 2008). Furthermore, long-term association with medicinal compounds might lead to relaxed selection on immune genes. The genome of honey bees (*Apis mellifera*) has a reduced number of canonical insect immune genes, possibly due to the use of medicinal compounds and behavioral defense mechanisms (Evans et al., 2006). Our results show that all four significantly differentially expressed canonical immune genes were down-regulated in monarchs fed with *A. curassavica*, which is in line with the hypothesis that medicinal milkweeds lead to reduced investment in immunity.

Interestingly, one of the immune genes that was down-regulated in larvae feeding on *A. curassavica* is a FREP-like receptor (DPOGS203317). Previous studies of infection of insects by another apicomplexan parasite (*Plasmodium* in *Anopheles gambiae*), which also infects insects through the midgut wall, have shown that several fibrinogen-related proteins (FREPs) play an important role in anti-parasitic defense. For example, overexpression of FREP13 results in increased resistance to *Plasmodium* infection (Dong & Dimopoulos, 2009; Simões et al., 2017). In

contrast, inactivation of FREP1 increases resistance, because FREP1 functions as an important host factor that mediates *Plasmodium* ookinete's invasion of the mosquito midgut epithelium (Dong, Simões, Marois, & Dimopoulos, 2018; Zhang et al., 2015). Our results show down-regulation of a FREP-like gene when larvae feed on a milkweed that confers stronger resistance to parasite infection. However, the exact function of this FREP-like gene remains unknown. In addition, two other immune genes that were down-regulated when feeding on *A. curassavica* are CLIP serine proteases (DPOGS215180 and DPOGS213841). CLIP serine proteases are a large gene family (Christophides et al., 2002), and some of them play an important role in anti-malaria defense (Barillas-Mury, 2007; Volz, Müller, Zdanowicz, Kafatos, & Osta, 2006). Future studies that directly examine the function of these particular immune genes are needed to understand their potential role in defense against *O. elektroscirrha* infections.

3.4.4 Transcriptional responses in relation to parasite infection

Our study confirmed previous findings that monarch larvae fed with *A curassavica* (high-cardenolide) have stronger anti-parasite resistance than those fed with *A. incarnata* (low-cardenolide) (Fig. 3.1B). Nevertheless, we observed almost no transcriptional response to parasite infection regardless of host plant diet. There are three possible explanations for these results. First, the parasite might be able to suppress or evade the host immune system, which has been demonstrated in several other specialist parasites (Gurung & Kanneganti, 2015; MacGregor, Szöőr, Savill, & Matthews, 2012; Selkirk, Bundy, Smith, Anderson, & Maizels, 2003). Second, the infection may not induce a systemic response; the immune responses may instead have occurred locally and hence may not have been detectable when sequencing the transcriptome of the gut or body. Third, we chose a 24-hr timepoint post infection to try to capture host responses against

parasites invading into the body cavity, which is the period in the infection cycle when mosquitoes exhibit up-regulation in midgut-based immune responses to apicomplexan parasites (Blumberg, Trop, Das, & Dimopoulos, 2013; Vlachou, Schlegelmilch, Christophides, & Kafatos, 2005). However, it is possible that the parasite is more active and/or has a stronger interaction with the host immune system at different stages of the infection cycle. Thus, additional life stages should be taken into consideration in future analyses.

3.5 Conclusions

We compared transcriptional profiles of monarch larvae fed two different milkweed species and examined larval transcriptional responses to infection by a specialist parasite. Our results demonstrate that monarch larvae differentially express hundreds of genes when feeding on *A. curassavica* or *A. incarnata*, two milkweed species that differ strongly in their secondary chemical content. Those differentially expressed genes include genes within multiple families of canonical insect detoxification genes, suggesting that they may play a role in processing plant diets with different levels of toxicity. Notably, all ABC transporters were up-regulated in monarchs fed with *A. incarnata*, the less toxic plant, which might be related to an increasing level of cardenolide sequestration. Interestingly, the few immune genes that were differentially expressed in monarchs reared on the two plant species were all down-regulated on the anti-parasitic *A. curassavica*, consistent with the hypothesis that medicinal plants could reduce immune investment by providing an alternative form of anti-parasite defense.

3.6 Acknowledgments

We thank W. Palmer for providing immune gene annotation resources for *Danaus plexippus*, A. Berasategui, E. Diaz-Almeyda, H. Salem, C. Beck, S. Birnbaum, B.-W. Lo, E. Whittington, and V. Talla for helpful discussion on this research, A. Salmi and K. Baffour-Ado for help with the experiments, and H. Streit for performing the chemical analyses. Computational analyses were performed on resources provided by the University of Kansas Information and Telecommunication Technology Center. This work was supported by National Science Foundation (NSF) grant IOS-1557724 to JCdR, NMG, and MDH, NSF Graduate Research Fellowship Program DGE-1444932 to EVH, and NSF grant DEB-1457758 to JRW.

3.7 Supplemental Information

Table S3.1. Sample sizes for each RNA-Seq. treatment group.

Hostplant species	Infection treatment	Tissue type	Sample size
A. incarnata	Infected	Gut	4
A. incarnata	Infected	Body	4
A. incarnata	Uninfected	Gut	3
A. incarnata	Uninfected	Body	3
A. curassavica	Infected	Gut	3
A. curassavica	Infected	Body	3
A. curassavica	Uninfected	Gut	3
A. curassavica	Uninfected	Body	3

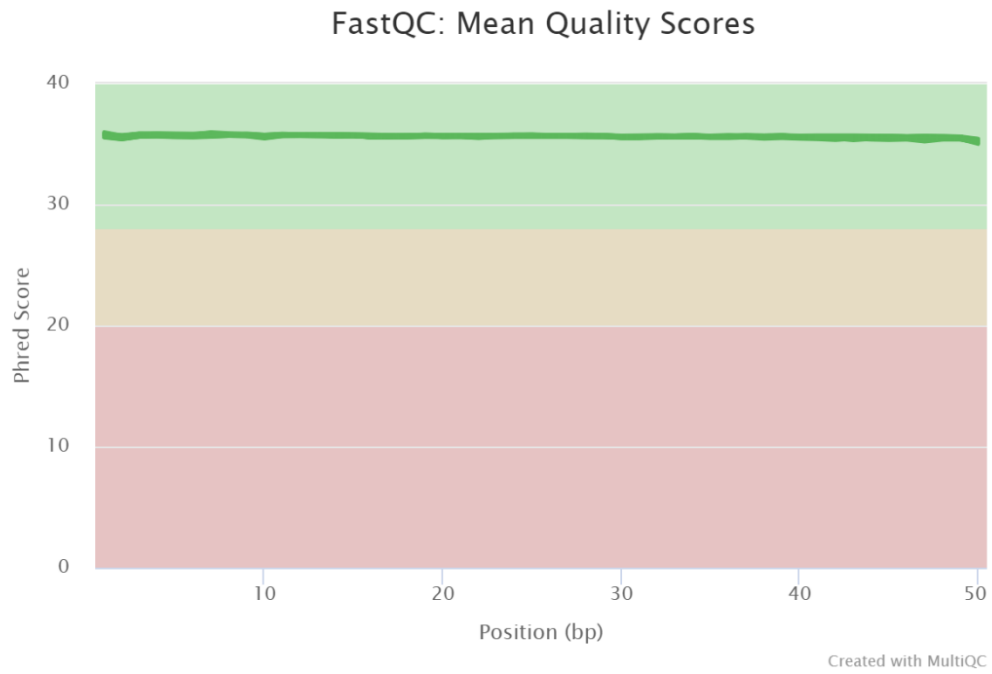
Table S3.2. List of immune genes used in this study.

Gene ID	Gene name	Gene length	Functional class	Chromosome
DPOGS200905	PGRP-like	3371	Recognition	Z-chromosome
DPOGS209813	PGRP-like	1574	Recognition	Autosome
DPOGS206909	PGRP-like	1472	Recognition	Z-chromosome
DPOGS206910	PGRP-like	7919	Recognition	Z-chromosome
DPOGS207148	PGRP-like	15320	Recognition	Z-chromosome
DPOGS209814	PGRP-like	2222	Recognition	Autosome
DPOGS206026	PGRP-like	3333	Recognition	Z-chromosome
DPOGS212963	BGRP-like	7393	Recognition	Autosome
DPOGS215599	BGRP-like	2910	Recognition	Autosome
DPOGS212940	BGRP-like	4002	Recognition	Autosome
DPOGS212941	BGRP-like	3544	Recognition	Autosome
DPOGS212964	BGRP-like	3508	Recognition	Autosome
DPOGS212965	BGRP-like	3969	Recognition	Autosome
DPOGS203317	Frep-like	5328	Recognition	Z-chromosome
DPOGS206045	Frep-like	4504	Recognition	Z-chromosome
DPOGS203951	Frep-like	10862	Recognition	Z-chromosome
DPOGS210549	Class B-like SCR	5148	Recognition	Autosome
DPOGS203180	Class B-like SCR	13541	Recognition	Autosome
DPOGS202796	Class B-like SCR	9956	Recognition	Autosome
DPOGS214397	Other SCR	13196	Recognition	Autosome
DPOGS213636	Other SCR	15159	Recognition	Autosome
DPOGS212634	Other SCR	14283	Recognition	Autosome
DPOGS202826	Other SCR	9352	Recognition	Autosome
DPOGS215836	TEP-like	10955	Recognition	Autosome
DPOGS210251	NIM-like	13838	Recognition	Autosome
DPOGS210210	NIM-like	5318	Recognition	Autosome
DPOGS210211	NIM-like	13481	Recognition	Autosome
DPOGS204835	CLIP-like	3824	Modulation	Autosome
DPOGS205231	CLIP-like	27491	Modulation	Autosome
DPOGS206561	CLIP-like	4655	Modulation	Autosome
DPOGS215180	CLIP-like	11215	Modulation	Autosome
DPOGS215181	CLIP-like	10878	Modulation	Autosome
DPOGS206562	CLIP-like	14594	Modulation	Autosome
DPOGS206563	CLIP-like	3009	Modulation	Autosome
DPOGS213841	CLIP-like	3609	Modulation	Autosome
DPOGS215183	CLIP-like	13397	Modulation	Autosome
DPOGS215188	CLIP-like	4339	Modulation	Autosome

Gene ID	Gene name	Gene length	Functional class	Chromosome
DPOGS204146	CLIP-like	5061	Modulation	Autosome
DPOGS204147	CLIP-like	6895	Modulation	Autosome
DPOGS201678	CLIP-like	3492	Modulation	Autosome
DPOGS215220	CLIP-like	3818	Modulation	Autosome
DPOGS215098	CLIP-like	27161	Modulation	Autosome
DPOGS208169	CLIP-like	3911	Modulation	Autosome
DPOGS201966	CLIP-like	10337	Modulation	Autosome
DPOGS215182	CLIP-like	10004	Modulation	Autosome
DPOGS211355	CLIP-like	3601	Modulation	Autosome
DPOGS203664	CLIP-like	7166	Modulation	Autosome
DPOGS210568	CLIP-like	8283	Modulation	Autosome
DPOGS214570	CLIP-like	4377	Modulation	Autosome
DPOGS204148	CLIP-like	2905	Modulation	Autosome
DPOGS205210	CLIP-like	3342	Modulation	Autosome
DPOGS211237	CLIP-like	2943	Modulation	Autosome
DPOGS206224	CLIP-like	5720	Modulation	Autosome
DPOGS206217	CLIP-like	3300	Modulation	Autosome
DPOGS205206	CLIP-like	3310	Modulation	Autosome
DPOGS209809	SPZ-like	1777	Signaling - Toll	Autosome
DPOGS209810	SPZ-like	5018	Signaling - Toll	Autosome
DPOGS203200	Toll_like-receptors	3821	Signaling - Toll	Autosome
DPOGS205279	Toll_like-receptors	4664	Signaling - Toll	Autosome
DPOGS202626	Toll_like-receptors	2709	Signaling - Toll	Autosome
DPOGS205281	Toll_like-receptors	3894	Signaling - Toll	Autosome
DPOGS205295	Toll_like-receptors	3887	Signaling - Toll	Autosome
DPOGS205123	Toll_like-receptors	5897	Signaling - Toll	Autosome
DPOGS211472	Toll_like-receptors	4533	Signaling - Toll	Autosome
DPOGS200002	Toll_like-receptors	5428	Signaling - Toll	Autosome
DPOGS205283	Toll_like-receptors	1949	Signaling - Toll	Autosome
DPOGS215274	Toll_like-receptors	5667	Signaling - Toll	Autosome
DPOGS203198	Toll_like-receptors	3410	Signaling - Toll	Autosome
DPOGS205293	Toll_like-receptors	869	Signaling - Toll	Autosome
DPOGS205296	Toll_like-receptors	2006	Signaling - Toll	Autosome
DPOGS207788	Tollip	3828	Signaling - Toll	Autosome
DPOGS205936	MyD88	2965	Signaling - Toll	Autosome
DPOGS208945	Tube	2121	Signaling - Toll	Autosome
DPOGS214647	Pellino	3601	Signaling - Toll	Autosome
DPOGS210260	Pelle	11577	Signaling - Toll	Autosome

Gene ID	Gene name	Gene length	Functional class	Chromosome
DPOGS202662	TRAF2	6521	Signaling - Toll	Autosome
DPOGS209243	ECSIT	1486	Signaling - Toll	Autosome
DPOGS209453	Cactus	2181	Signaling - Toll	Autosome
DPOGS215778	IMD	1319	Signaling - IMD	Autosome
DPOGS200403	TAK1	12371	Signaling - IMD	Autosome
DPOGS202907	IKKgamma	14063	Signaling - IMD	Autosome
DPOGS202564	IKKbeta	1586	Signaling - IMD	Autosome
DPOGS207960	FADD	916	Signaling - IMD	Autosome
DPOGS212093	Dredd	1508	Signaling - IMD	Autosome
DPOGS200977	Tab2	3087	Signaling - IMD	Autosome
DPOGS203759	IAP2	5539	Signaling - IMD	Autosome
DPOGS201405	Ubc13	455	Signaling - IMD	Autosome
DPOGS208954	Hem	3377	Signaling - JNK	Autosome
DPOGS213169	JNK	3769	Signaling - JNK	Autosome
DPOGS214573	Fos	2706	Signaling - JNK	Autosome
DPOGS202887	Jun	242	Signaling - JNK	Autosome
DPOGS214325	PIAS	12079	Signaling - JAK-STAT	Autosome
DPOGS214451	SOCS	2617	Signaling - JAK-STAT	Autosome
DPOGS200349	DOMELESS	2960	Signaling - JAK-STAT	Autosome
DPOGS210157	Hopscotch	8831	Signaling - JAK-STAT	Autosome
DPOGS212956	Stat	16032	Signaling - JAK-STAT	Autosome
DPOGS213997	Attacin-Like	877	Effector	Autosome
DPOGS205720	Attacin-Like	1439	Effector	Autosome
DPOGS215451	Attacin-Like	818	Effector	Autosome
DPOGS210270	Cecropin-like	374	Effector	Autosome
DPOGS210268	Cecropin-like	422	Effector	Autosome
DPOGS210269	Cecropin-like	529	Effector	Autosome
DPOGS200256	Cecropin-like	428	Effector	Autosome
DPOGS210271	Cecropin-like	352	Effector	Autosome
DPOGS210265	Cecropin-like	848	Effector	Autosome
DPOGS210304	Gloverin-like	644	Effector	Autosome
DPOGS210303	Gloverin-like	822	Effector	Autosome
DPOGS202093	NOS-like	15626	Effector	Autosome
DPOGS202094	NOS-like	19723	Effector	Autosome
DPOGS201818	PPO-like	4660	Effector	Z-chromosome
DPOGS201819	PPO-like	5674	Effector	Z-chromosome
DPOGS206820	PPO-like	9352	Effector	Z-chromosome
DPOGS200017	PPO-like	5087	Effector	Autosome

(A)



(B)

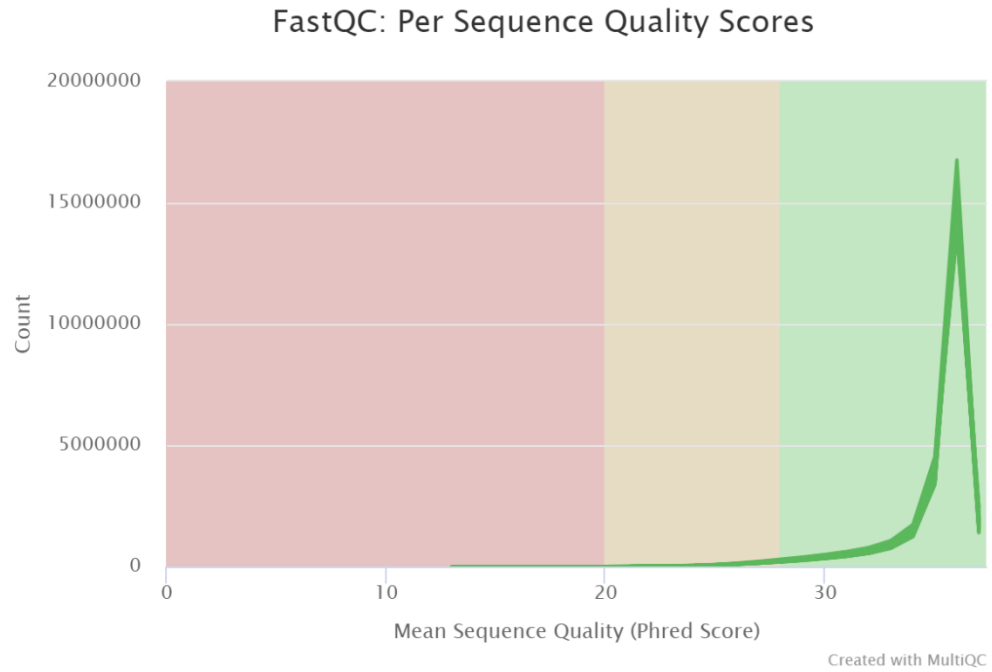


Figure S3.1. Quality of RNA-Seq reads across all samples. (A) The mean quality score across each base position in the read. (B) The number of reads with each mean quality score. Each green curve represents one RNA-Seq sample (N = 26). The curves are highly overlapped across samples because of highly consistent quality.

Chapter 4: Population genomics reveals complex patterns of immune gene evolution in monarch butterflies (*Danaus plexippus*)

Wen-Hao Tan, Andrew J. Mongue, Jacobus C. de Roode, Nicole M. Gerardo, James R. Walters

Abstract

Immunity-related genes presumably rapidly evolve as parasites and pathogens exert strong selection pressures on host defense, but the evolution of immune genes is also constrained by trade-offs with other biological functions and shaped by the environmental context. Thus, immune genes may exhibit complex evolutionary patterns, particularly when organisms disperse to or live in variable environments. In this study, we examine the evolutionary patterns of the full set of known canonical immune genes within and among populations of monarch butterflies (*Danaus plexippus*), and relative to a closely related species (*Danaus gilippus*). Monarchs represent a system with a known evolutionary history, in which North American monarchs dispersed to form novel populations elsewhere, providing an opportunity to explore the evolution of immunity in the light of population expansion into novel environments. We found that immune genes as a whole do not exhibit consistent patterns of selection, differentiation, or genetic variation, but that patterns are tied to functional classes. Specifically, species comparisons between *D. plexippus* and *D. gilippus* and analyses of monarch populations both revealed consistently low levels of genetic variation in signaling genes, suggesting conservation of these genes over evolutionary time. Modulation genes showed the opposite pattern, with signatures of relaxed selection across populations. In contrast, recognition and effector genes exhibited less consistent patterns. When focusing on outlier genes

(those with exceptionally strong signatures of selection or differentiation), we also found population-specific patterns, consistent with the hypothesis that monarch populations do not face uniform selection pressures with respect to immune function.

4.1 Introduction

The cellular and humoral immune systems provide one of the primary animal defenses against parasites and pathogens. Given that parasites and pathogens exert strong selection pressure on their hosts, immunity-related genes are presumed to be under selection and rapidly evolving due to host-pathogen coevolutionary arms races (McTaggart et al., 2012; Schlenke & Begun, 2003). However, the evolution of immune genes is also constrained by trade-offs with other biological functions and shaped by environmental context (Demas & Nelson, 2012). When animals colonize novel environments, they often encounter novel ecological conditions, including diets, pathogens, and microbes, that could influence disease susceptibility and alter selection pressures on immune functions (Eizaguirre, Lenz, Kalbe, & Milinski, 2012). In addition to their canonical cellular and humoral immune defenses, animals may use medicinal compounds or symbionts to protect against parasites and pathogens (Parker et al., 2011). Utilization of alternative defenses may vary across populations due to environmental context, selection, plasticity, and drift. These differences, in turn, could shape the evolution of immune genes across populations. Taken together, the evolutionary patterns of immune genes may be complicated, particularly when organisms disperse to novel environments.

The cellular and humoral immune system of insects is relatively simple compared to the immune system in vertebrates, potentially facilitating the study of immune gene evolution.

Previous work has shown that the canonical immune system of insects mainly consists of four functional classes: recognition (e.g., peptidoglycan recognition proteins or PGRPs), signaling (e.g., the Toll signaling pathway), modulation (e.g., CLIP serine proteases), and effector (e.g., antimicrobial peptides: AMPs) (Christophides et al., 2002). Insect immune responses usually begin with the identification of foreign molecules by pattern recognition receptors (PRRs) encoded by recognition genes. The recognition of foreign molecules activates downstream signaling cascades that involve proteins encoded by signaling and modulation genes. For instance, recognition of Gram-positive bacteria and fungi generally triggers the activation of the Toll pathway, while recognition of Gram-negative bacteria generally triggers the activation of the *immune deficiency* (IMD) pathway. These signaling cascades lead to the production of effector proteins (e.g., AMPs, pro-phenoloxidases that lead to melanization responses) that directly interact with pathogens and other invaders (Lemaitre & Hoffmann, 2007).

Some studies of insect immune gene evolution have demonstrated that immune genes are rapidly evolving. For example, Erler *et al.* (2014) showed that AMPs evolve much faster than non-immune genes in multiple bumblebee species, and Viljakainen *et al*. (2009) demonstrated that a select subset of immune genes (14 recognition and effector genes) are rapidly evolving in both honey bees and ants. However, these studies and most others have focused on only a few genes or one part of the immune system, without consideration of the full set of canonical immune genes.

Consideration of the immune gene set as a whole is important, in part, because different immune components may face different selection pressures. Specifically, coevolutionary theory would predict that molecules that directly interact with rapidly evolving pathogens and parasites – such as those encoded by recognition and effector genes – may undergo faster evolution than those involved in signal transduction. Indeed, a comparative study of twelve *Drosophila* species found

that recognition proteins and effectors are rapidly evolving and highly differentiated; in contrast, proteins within signaling transduction cascades are more constrained across species (Sackton et al., 2007).

Highlighted by the above-mentioned studies, most research on insect immune gene evolution has been conducted mostly on subsets of immune genes, and through comparisons across species. As such, we still lack a proper understanding of the evolution of immune genes across populations within species. To our knowledge, only a few studies have taken a comprehensive, population-centered approach: Early *et al.* (2017) and Keehnen *et al.* (2018) examined the evolution of the full set of canonical immune genes across populations in fruit flies (*Drosophila melanogaster*) and a butterfly (*Pieris napi*), respectively. Both studies demonstrated that immune gene functional classes vary in their patterns of selection and differentiation, with conservation of signaling genes, balancing selection acting on effector genes, and recognition genes showing higher levels of between-population differentiation.

In this study, we examined the evolution of the full set of canonical immune genes across natural populations of monarch butterflies (*Danaus plexippus*). Monarch butterflies are widely-distributed, specialist herbivores that feed on toxic milkweed plants during their larval stage (Ackery & Vane-Wright, 1984; Oberhauser & Solensky, 2004). Monarchs originated in North America and colonized worldwide locations in the 19th century through independent dispersal events across the Pacific Ocean, the Atlantic Ocean, and across Central-South American (Fig. 4.1) (Ackery & Vane-Wright, 1984; Zhan et al., 2011), providing an opportunity to study immune gene evolution in the context of a known evolutionary history. Importantly, through these dispersal events, monarchs formed populations in which they relied on more toxic milkweed host plants and in which they experienced greater risk of infection by the common monarch parasite *Ophryocystis*

elektroscirrha (Altizer & de Roode, 2015), likely altering selection on the monarch immune system. Here, we assessed patterns of divergence, diversity, and selection, for the full set of known, canonical immune genes identified in monarchs, using *D. gillipus* as an outgroup and contrasting the ancestral North American monarch population with geographically and genetically distinct derived populations. We compared immune genes to selected genomic controls, paired for gene length and physical location, to determine if immune genes differ from the genomic background for patterns of polymorphism and selection. Overall, we demonstrate that immune genes as a unit are not under uniform pattern of selection or differentiation, but that different components of the immune system exhibit distinct patterns. Furthermore, variable patterns across populations are consistent with the hypothesis that the worldwide dispersal of monarchs resulted in differential selection on immune genes.

Figure 4.1. Current distribution of monarch butterflies around the world and their historical dispersal routes. Monarchs originated in the North America and established other populations via three main dispersal events: across the Pacific Ocean, across the Atlantic Ocean, and toward

Central/South America.

4.2 Materials and Methods

4.2.1 Overview of approach

Differential selection pressures owing to ecological differences could affect the type and strength of selection on immune genes. In addition to selection, other factors such as demographic history and local genomic factors also may affect their evolutionary patterns. Given that several population genetic measures of selection are sensitive to demographic effects, past demographic history and recent dispersal are important factors that could influence and/or confound signatures of selection observed (Eyre-Walker & Keightley, 2009; Vitti, Grossman, & Sabeti, 2013). However, in most population genomic studies of immune genes, relatively little is known about the demographic history and the ecological differences of the focal populations. In monarch butterflies, previous population genetic and genomic studies have inferred that monarchs originated in North America and recently spread around the world via three major dispersal events (Pierce et al., 2014; Zhan et al., 2014). While these events led to formation of populations subject to different ecological conditions, the dispersal process itself may also influence patterns of population genetics. To account for this, we used a paired-control approach to determine if signatures of selection in functional classes of immune genes differ from those in the background genome. In addition, we identified individual immune genes that are genome-wide outliers for combinations of population genetic parameters, indicating they are likely experiencing exceptional selective pressures.

4.2.2 The population genomic dataset

We obtained a whole genome Illumina sequencing dataset from Zhan *et al.* (2014), which

sequenced monarch samples across populations worldwide. Based on previous population genetic and genomic studies (Pierce et al., 2014; Zhan et al., 2014), we assigned monarch samples into genetic populations according to their collection location. We excluded samples with average sequencing depth lower than 10X for quality control purposes. We used a total of 37 whole monarch genomes in our study, including the ancestral population (North America) and derived populations in South Florida, the Pacific, and the Atlantic (Fig. 4.1, supplemental information Table S4.1).

We aligned sequencing reads to the reference monarch genome (Zhan et al., 2011) using Bowtie2 with the option "--very-sensitive-local" (Langmead & Salzberg, 2012). After reference mapping, we indexed and sorted the BAM files, and removed PCR duplicates using Picard (http://broadinstitute.github.io/picard/). We performed local realignment using GATK (McKenna et al., 2010).

4.2.3 Gene sets

We obtained a full set of annotated monarch immune genes published by the Heliconius Genome Consortium (2012), which included a set of annotated (*Heliconius*) immune genes and their orthologs in several species, including monarchs. The monarch orthologs listed in this published dataset were based on a previous version of monarch genome annotation (OGS1.0), so we updated this full set of immune genes to the latest version of gene annotation (OGS2.0) using information provided in Monarch Base (Zhan & Reppert, 2013). This updated monarch immune gene set contains 114 genes belonging to functional classes of recognition, signaling, modulation, and effector (see supplemental information Table S4.2). We also obtained the latest version (OGS2.0) of all the annotated monarch genes from the published reference genome (Zhan et al., 2011; Zhan & Reppert, 2013) in order to compare evolution of immune genes to evolution of non-

immune genes (as controls) in the genome background.

We restricted our analyses to autosomal genes to avoid the complication of unequal sampling between autosomes and the Z sex chromosome; sequenced individuals were of different sexes so a variable number of Z chromosomes were sampled. We did not perform a separate analysis of Z-linked genes due to sample size limitations. We identified Z-linked immune genes based on chromosomal assignments obtained from (Mongue, Nguyen, Volenikova, & Walters, 2017). The majority of immune genes are on autosomes, with only 12 genes located on the Z chromosome (see supplemental information Table S4.2).

4.2.4 Population genetic analyses

We calculated four population genetic statistics: pairwise nucleotide diversity (π), Watterson's θ, Tajima's D, and F_{ST}. Nucleotide diversity (π) and Watterson's θ are two measures describing the degree of genetic diversity within a population, with greater values indicating greater diversity (Nei, 1979; Watterson, 1975). Tajima's D (Tajima, 1989), which is computed using the difference between these two measures of genetic diversity, provides information on signatures of selection or demographic events. A Tajima's D value close to 0 indicates neutrality. A more negative Tajima's D value represents an excess of low- or high-frequency polymorphisms, indicating directional selection or population expansion. A more positive Tajima's D value represents low levels of both low- and high-frequency polymorphisms, which indicates balancing selection or population contraction (Nielsen & Slatkin, 2013; Vitti et al., 2013). F_{ST} is a measure of population differentiation due to genetic structure, with a value ranging from 0 to 1. An F_{ST} value equal to zero indicates no differentiation. An F_{ST} value equal to one indicates that different alleles are fixed in different populations (Wright, 1921).

We generated folded site frequency spectra and calculated the four statistics using ANGSD (Korneliussen, Albrechtsen & Nielsen, 2014). We calculated pairwise nucleotide diversity (π), Watterson's θ and Tajima's D for each population; we calculated F_{ST} between populations by comparing each of the three derived populations (*i.e.*, Florida, Pacific, and Atlantic) to the ancestral population (North America). For all calculations, we first generated a site frequency spectrum (SFS) for all genes in the same functional class to use as a prior for gene-specific parameter estimates. Using this prior, we then calculated those four population genetic statistics for each gene in the functional class. We repeated the procedures for each gene with either: (1) zero-fold degeneracy sites; (2) four-fold degeneracy sites; and (3) all sites within each gene. The zero-fold and four-fold degeneracy sites for all monarch genes were obtained from Mongue *et al.* (2018). The genomic positions of each gene were obtained from the latest version of gene annotation (OGS2) in Monarch Base (Zhan et al., 2011). We performed all calculations for all genes in the genome. We generated inputs for ANGSD and processed the data using custom R and python scripts in R version 3.4.1 (R Core Team, 2017) and python version 2.7.5.

4.2.5 A paired control approach to compare immune genes to the genomic background

Evolutionary change of a gene can be influenced by gene length and local genomic factors, such as recombination rate and selection on nearby genes (Castellano, Coronado-Zamora, Campos, Barbadilla, & Eyre-Walker, 2016; Comeron, Ratnappan, & Bailin, 2012; Wong et al., 2008). Therefore, we evaluated whether immune genes differed from broader patterns in the genome background using a paired-control approach that compares immune genes to a selected subset of control genes. This paired-control approach enables us to take these factors into consideration, assessing the patterns of selection more conservatively; our approach is similar to that used by

Early et al. (2017) and Chapman, Hill, & Unckless (2018).

Specifically, we first constructed a pool of control genes for each immune gene based on the following criteria: (1) the length of the control genes are within either 0.5-2 times, or ±1500 bp, of the total length of the immune gene; (2) control genes are on the same scaffold (and thus chromosome) as the immune gene; (3) control genes are not known to have immune function. Given that a high proportion of scaffolds in the reference monarch genome are relatively small in size (N50 = 715 kbp) (Zhan & Reppert, 2013), in some cases control gene pools were small. When a candidate gene pool was smaller than eight genes, we relaxed the location criterion and expanded the search to the whole chromosome level, while keeping the other two criteria unchanged. In all cases, we were able to gather > 8 candidate genes. Four focal immune genes did not have a chromosomal assignment. For these, we searched for genes that also did not have chromosomal assignments that fit the size and gene function criteria. We excluded genes that did not have an adequate number of 0-fold or 4-fold sites for estimating population genetic statistics from the control gene pools. For a given immune functional group, we calculated the test statistic as the summation of the difference between an immune gene and the mean of its control genes. We determined significance through 10,000 permutations. For each permutation round, we randomly sampled one gene for each immune gene from a pool containing the immune gene itself and all corresponding control genes with replacement to serve as the test gene, and calculate the difference between the test gene and the mean of the remaining genes in the pool. The permuted test statistic is calculated as the summation of those differences for genes belonging to a given immune functional group. We calculated *P*-values as the percentage of the 10,000 permutations in which the absolute value of the test statistic (observed value) is less than the absolute mean value of the permuted sets (permuted null distribution). The paired-control analyses were performed using

custom R scripts in R version 3.4.1 (R Core Team, 2017).

4.2.6 Between-species analyses

In addition to between-population comparisons, we also sought to estimate longer-term evolutionary patterns by leveraging whole genome sequencing of a congener, the queen butterfly (*Danaus gilippus*) (Zhan et al., 2014). This gave us the opportunity to look at scaled rates of divergence between species (Dn/Ds). We aligned *D. gilippus* reads to the monarch reference using the stampy alignment software (Lunter & Goodson, 2011), parameterized for an increased (5%) substitution rate between reads and reference. These data were then taken through GATK's best practice pipeline for SNP calling, including quality filtering of variants (McKenna et al., 2010). Passing variants were classified as synonymous or non-synonymous by SNPeff (Cingolani et al., 2012). Finally, we calculated Dn per gene as the number of nonsynonymous substitutions per non-synonymous site (and likewise for Ds), using previous knowledge of the degeneracy of each position in a coding sequence (Mongue et al., 2019). We used R to perform Mann-Whitney U tests to assess significance of differences in divergence rates for immune gene classes compared to the control gene sets described above.

The *D. gillipus* data additionally allowed us to estimate the proportion of substitutions driven by adaptation (α) for immune genes and to compare with estimates from corresponding control genes in the monarch genome. As with established methods (Mongue et al., 2019), we used the queen butterfly sequences to infer a parsimonious ancestral (allele) state at polymorphic sites in the monarch genome, allowing us to generate an unfolded site frequency spectrum, *i.e.* one that differentiates ancestral and derived allele frequencies. With unfolded spectra, we employed the likelihood model implemented in polyDFE (Tataru, Mollion, Glémin, & Bataillon, 2017) to

estimate α and the distribution of fitness effects of new non-synonymous mutations (DFEs) while accounting for demography and errors in allele frequency polarization. To assess uncertainty in these estimates, parametric bootstrapping of input site frequency spectra (as implemented in polyDFE) was used to obtain a distribution of α and DFE statistics. Differences in α were tested via another series of Mann Whitney U tests. Differences between DFEs were not formally tested but were used as ancillary, qualitative inferences to contextualize related results. Bootstrapping, statistical analyses, and visualization were completed with custom R scripts.

4.2.7 Outlier analyses

To identify specific loci which may experience distinctive evolutionary pressures, we searched for immune genes which are outliers relative to the genome-wide distributions of population genetic parameters. While genome-wide patterns will be shaped by population demography, outliers likely reflect non-neutral processes. In particular, we jointly considered Tajima's D and Fst, reasoning that loci showing extreme values for both parameters are likely to be of particular interest. We performed the analyses across all genes in the genome at either 0-fold sites or 4-fold sites and used information at genome-wide 0-fold or 4-fold sites as prior for estimating SFS in ANGSD. We converted Tajima's D and F_{ST} values into percentiles in their genome-wide distribution. We defined genes that were either in the $< 2.5^{th}$ percentile or $>97.5^{th}$ percentile as genome outliers. To assess the outlier patterns considering both selection and population differentiation, we evaluated the relationship of Tajima's D and F_{ST} for each functional class. Converting the values into percentiles also enabled us to compare patterns across populations. We visualized the patterns by plotting the Tajima's D and F_{ST} genome percentiles against each other in a 2-D plot with Tajima's D on the x-axis and F_{ST} on the y-axis. Separating the plot by the genome median of the two measures, it

contains four areas: top-right ($x > 0.5$ & $y > 0.5$), bottom-right ($x > 0.5$ & $y < 0.5$), top-left ($x < 0.5$ and $y > 0.5$), and bottom-left ($x < 0.5$ and $y < 0.5$). Outliers falling into each of the four areas suggest different evolutionary scenarios: "top-right" suggests balancing selection acting differently between populations, "bottom-right" suggests balancing selection acting similarly between populations, "top-left" suggests directional selection acting differently between populations, and "bottom-left" suggests directional selection acting similarly between populations. We summarized the number of outliers in each area in contingency tables and analyzed the patterns. Due to small count numbers in some cells, we used Fisher's exact tests. In addition, we examined whether immune genes are disproportionally represented in genome-wide outliers using Chi-square tests. All statistical analyses were performed in R version 3.4.1.

4.3 Results

4.3.1 North America: the ancestral population

A. Within-species analyses: characterizing genetic diversity and signatures of selection

We measured the genetic diversity of immune genes using both pairwise nucleotide diversity (π) and Watterson's θ (Table 4.1 and Fig. 4.2). As a group, immune genes showed slightly lower genetic diversity compared to paired-control genes, though this result was not statistically significant. However, levels of genetic variation varied notably among the different functional classes of immune genes. At 0-fold sites, recognition and modulation genes exhibited a trend toward higher genetic variation than their respective control genes, while signaling and effector genes showed a trend toward lower genetic variation than their respective control genes. Signaling genes had a significantly lower π and Watterson's θ at the 0-fold sites than controls, while other

functional groups did not differ significantly from their controls. At 4-fold sites, none of the functional groups differed significantly from their controls; only the signaling genes had a marginally significantly lower π compared to controls.

We measured signatures of selection using Tajima's D (Table 4.1 and Fig. 4.2). Immune genes as a whole did not show a distinct pattern of selection; the full set of immune genes was not significantly different from the paired-controls. However, as with π and Watterson's θ, patterns of Tajima's D varied across different functional classes of immune genes. Recognition genes showed a trend of lower Tajima's D at both 0-fold and 4-fold sites but was only significantly lower at the 0-fold sites. Signaling genes showed a significantly lower Tajima's D than controls at only the 4-fold sites. Modulation genes did not exhibit any significant differences to the controls. Effector genes showed significantly higher Tajima's D at the 4-fold sites and marginally significantly higher Tajima's D at the 0-fold sites to their respective controls.

Taken together, the full set of immune genes did not differ from control genes in either genetic diversity or signatures of selection; however, different functional classes exhibited significant differences. Specifically, signaling genes showed lower genetic variation than control genes, consistent with broad purifying selection; associated background selection could explain the reduced 4-fold site Tajima's D. By contrast, the strongly elevated Tajima's D among effector genes seems best explained by frequent balancing selection among these loci. Analyses based on all sites within each gene showed similar qualitative results (see supplemental information Table S3 and Fig. S1).

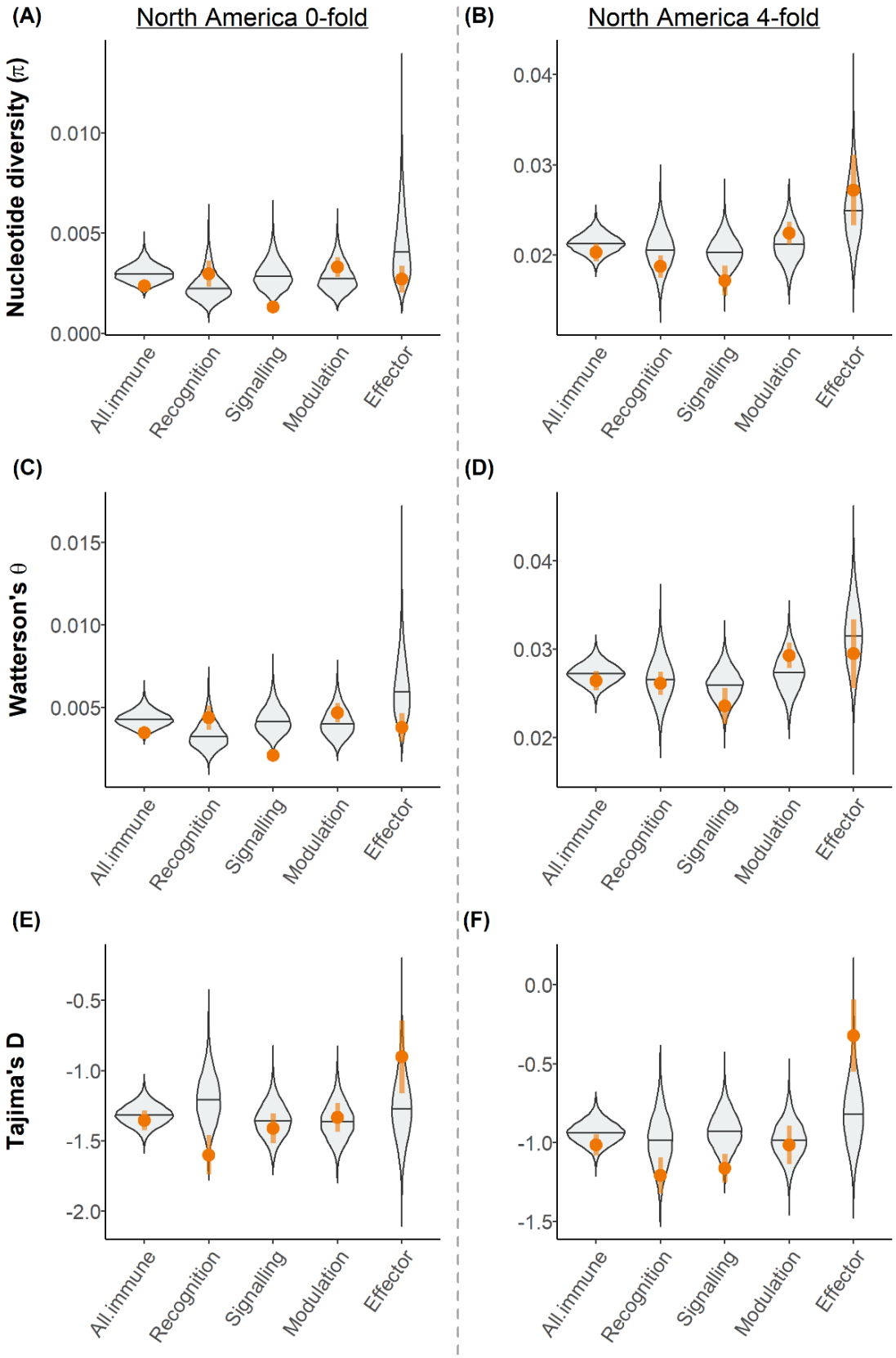

(A)
North America 0-fold
Nucleotide diversity (π)
0.010
0.005
0.000
(B)
North America 4-fold
0.04
0.03
0.02
(C)
Watterson's θ
0.015
0.010
0.005
(D)
0.04
0.03
0.02
(E)
Tajima's D
-0.5
-1.0
-1.5
-2.0
(F)
0.0
-0.5
-1.0
-1.5
All.immune
Recognition
Signalling
Modulation
Effector

Figure 4.2. Population genetic statistics of immune genes in the North American population using the paired-control approach. 0-fold sites shown in (A), (C), (E) and 4-fold sites shown in (B), (D), (F). (A)-(B): Nucleotide diversity ($\boldsymbol{\pi}$); (C)-(D): Watterson's θ; (E)- (F): Tajima's D. Each immune gene group was compared to selected pair-control sets. Violin plots show the distribution of the mean of each control set generated with 10, 000 permutations. The orange dots and vertical lines indicate mean ±1 SEM of the immune gene group of interest.

Table 4.1. Population genetic statistics of immune genes in the North American population using the paired-control approach. The upper half shows results based on the 0-fold sites and the lower half shows results based on the 4-fold sites. The F_{ST} section is non-applicable because the North American population was the reference population used for population comparisons. "All immune" indicates the full immune gene set. In each statistic, the first row shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % are labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 are labeled in bold. Asterisks indicate: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
0-fold sites					
π: test statistic	-0.07	0.01	-0.07	0.01	-0.02
π: > 0 (%)	5.84	83.10	**0.04**	75.68	14.60
π: *P*-value.	0.140	0.340	**0.024***	0.508	0.311
Watterson's θ: test statistic	-0.09	0.02	-0.09	0.02	-0.04
Watterson's θ: > 0 (%)	3.71	88.83	**0.03**	76.02	6.53
Watterson's θ: *P*-value	0.092	0.184	**0.012***	0.497	0.168
Tajima's D: test statistic	-4.25	-8.01	-2.38	0.69	5.44
Tajima's D: > 0 (%)	30.29	**0.76**	33.12	58.16	94.66
Tajima's D: *P*-value	0.599	**0.024***	0.643	0.859	0.098
F_{ST}: test statistic	NA	NA	NA	NA	NA
F_{ST}: > 0 (%)	NA	NA	NA	NA	NA
F_{ST}: *P*-value	NA	NA	NA	NA	NA
4-fold sites					
π: test statistic	-0.11	-0.04	-0.14	0.03	0.03
π: > 0 (%)	16.18	21.16	2.55	72.58	71.43
π: *P*-value.	0.327	0.419	0.059	0.564	0.596
Watterson's θ: test statistic	-0.09	-0.01	-0.10	0.06	-0.04
Watterson's θ: > 0 (%)	23.48	44.04	9.57	80.66	25.65
Watterson's θ: *P*-value	0.467	0.856	0.192	0.387	0.510
Tajima's D: test statistic	-8.60	-4.41	-10.33	-1.07	7.21
Tajima's D: > 0 (%)	13.10	9.24	**1.87**	38.65	**98.87**
Tajima's D: *P*-value	0.262	0.189	**0.043***	0.760	**0.015***
F_{ST}: test statistic	NA	NA	NA	NA	NA
F_{ST}: > 0 (%)	NA	NA	NA	NA	NA
F_{ST}: *P*-value	NA	NA	NA	NA	NA

B. Between-species analyses: comparing *D. plexippus* and *D. gilippus*

We further assessed molecular evolutionary patterns of immune genes by estimating divergence to the closely related queen butterfly. We tested for differences in rates of divergence (Dn/Ds) between immune genes and their controls selected from the rest of the genome. We found that neither effector (Fig. 3; W = 2866, $P = 0.764$) nor signaling genes (Fig. 3; W = 23427, $P = 0.352$) showed increased divergence compared to their controls, which is consistent with balancing and purifying selection respectively decreasing the fixation rate of variants. In contrast, we found elevated divergence in both modulation (Fig. 3; W = 6036.5, $P = 0.009$) and recognition genes (Fig. 3; W = 1156, $P = 0.018$) compared to their controls. Such a result is indicative of either increased directional selection or relaxed constraint allowing more non-synonymous differences to reach fixation. Taken together with within-species analyses of nucleotide diversity, these results suggest that relaxed selection is more likely for modulation genes, but the cause of increased divergence in recognition genes is less immediately apparent.

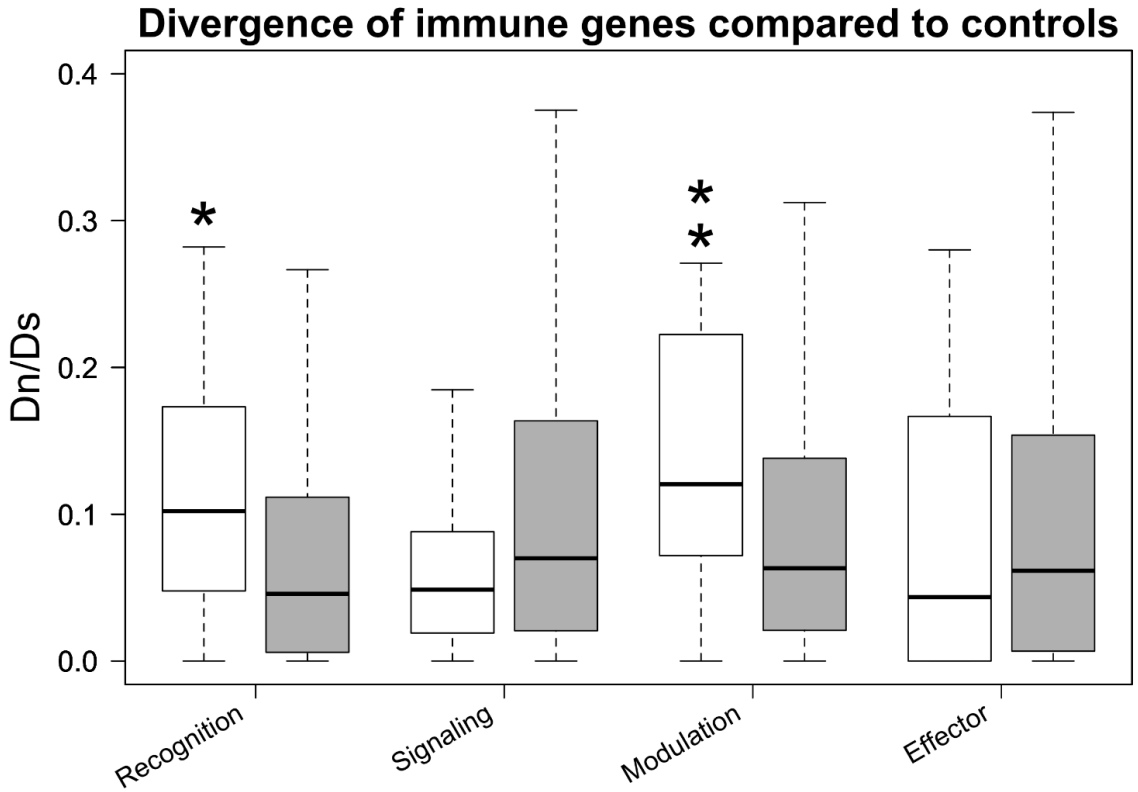


Figure 4.3. Divergence rates compared for immune genes and paired-controls using the queen butterfly as a reference. Here Dn is calculated as non-synonymous substitutions per non-synonymous site, and Ds is the number of synonymous substitutions per synonymous site. Rates for each gene class are labeled, with the control group in grey immediately to right. Asterisks represent levels of significance in a Mann-Whitney-U Test following the convention: * for <0.0.5 and ** < 0.01.

C. Distributions of fitness effects and estimates of adaptive evolution

To further investigate patterns of selection, we used SFS to estimate the distribution of fitness effects for new non-synonymous mutations (DFEs) among the immune gene functional classes and their control sets. Though we are unable to statistically compare differences between immune gene groups and their controls, the patterns are largely consistent with the results of other tests. Signaling genes exhibited a lack of neutral and weakly selected variants, combined with an increase in strongly deleterious and (to a lesser degree) beneficial variants (Fig. 4, second row). This pattern suggests most new variation is destined to be removed by purifying selection, with occasional

adaptive fixations. Modulation genes did not greatly differ from their control set, though the slight increase in inferred neutral and weakly selected variants ($-10 < s < 1$) is consistent with relaxed selection in this class of genes (Fig. 4, third row). Effector genes, however, showed a lack of strongly deleterious ($s < -100$) and an increase in moderately deleterious ($-100 < s < -10$) variants (Fig. 4, fourth row). This dearth of strongly deleterious variants suggests that alleles can reach more intermediate frequency, as expected under balancing selection.

Unlike the other classes of immune genes, the DFEs for recognition genes and their controls suggest an alternative explanation for the patterns observed in other population genetic statistics. Note that here, the control genes (Fig. 4, top right) exhibit a similar pattern to the one described above for the focal set of signaling genes, *i.e.* purifying selection. The recognition genes' DFEs, however, do not appear to be skewed by strong selection. In this light, other results for recognition genes may have more to do with purifying selection on controls than on selection on the recognition genes themselves.

Finally, we used the DFEs to estimate α (the proportion of adaptive substitutions) in each immune class and its control set. We found that α was significantly different between immune genes and controls in each of the four groups, as evidenced by non-overlapping confidence intervals (Fig. 5). For three of the four classes, the direction of these differences is consistent with other lines of evidence for selection. Namely, we found more adaptive evolution in effector and signaling genes and less adaptation in modulation genes compared to their controls. For recognition, however, evidence for less adaptation than controls conflicts with the evidence for selection from Tajima's D. This lower α, alongside the DFEs, suggest that recognition genes are under weaker selection than their paired controls.

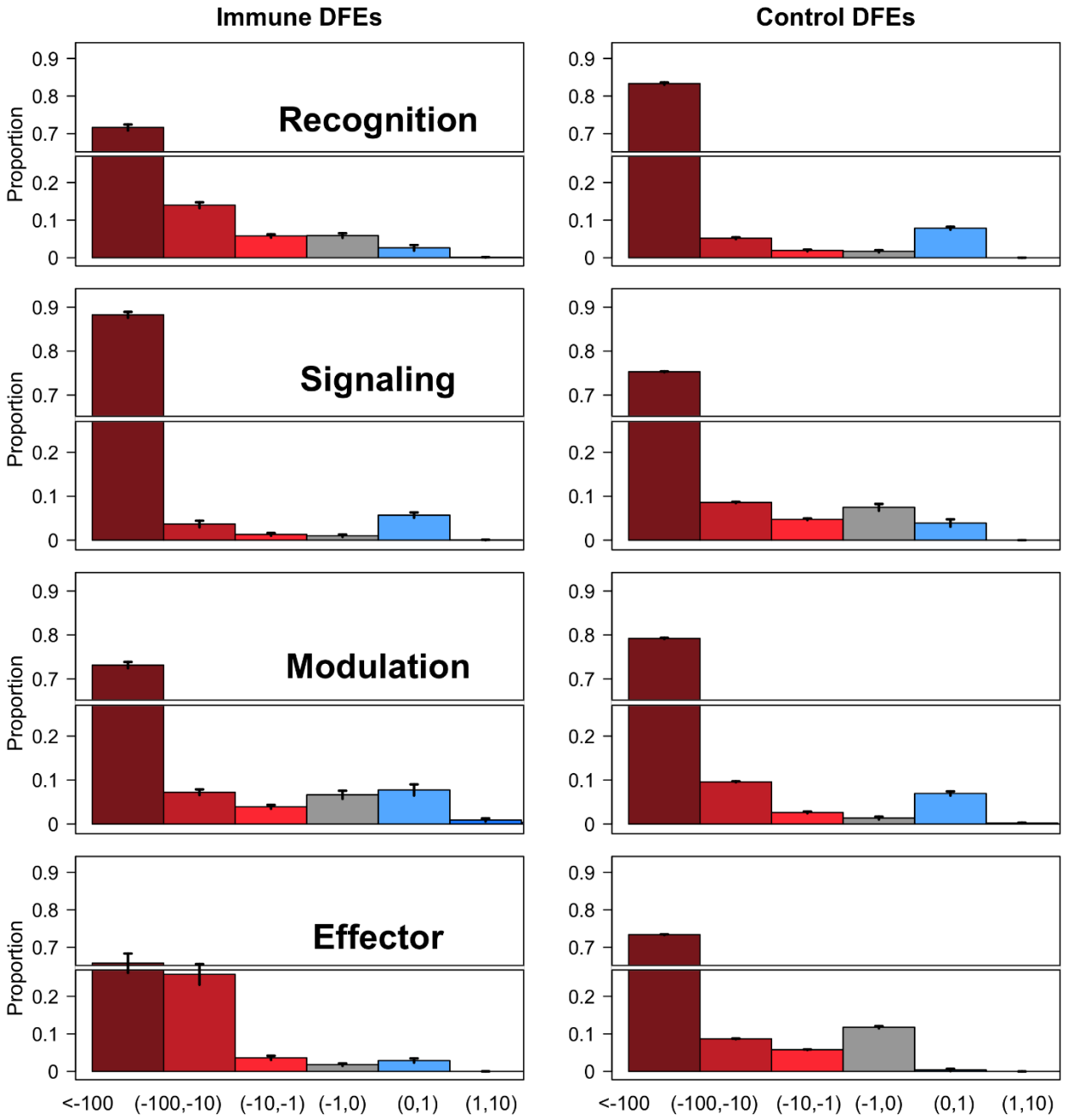


Figure 4.4. Predicted distributions of fitness effects of new, non-synonymous mutations for each of the four classes of immune genes (left column) and their paired-control sets (right column). Bars represent the proportion of variants that fall within a given selective class (s), from strongly deleterious (far left, darkest red) to beneficial (right, blue). Each plot is scaled with the same y-axis and has a gap from 0.25 to 0.65 to allow visualization of the whole distribution. Vertical lines on each bar, while mostly too small to notice, represent twice the standard error of the mean per-selective-class estimate from one hundred parametric bootstrap replicates. Estimates come from the tool polyDFE.

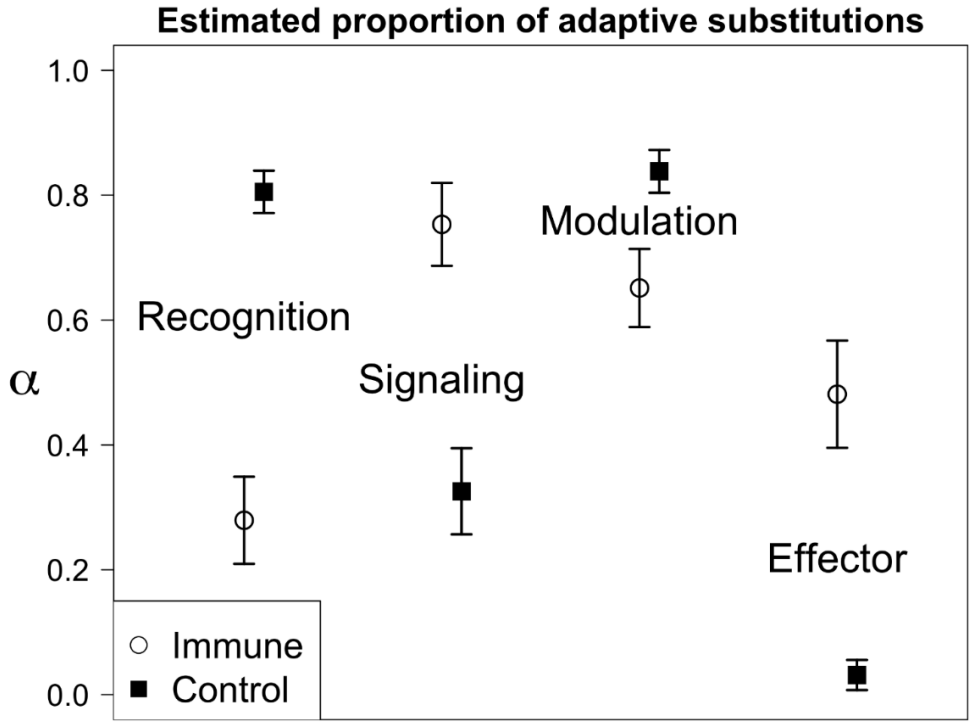


Figure 4.5. Estimates of the proportion of substitutions resulting from adaptive processes (α) based on DFEs computed in polyDFE. Each immune gene class (open circles) has a paired-control set of genes immediately to its right (filled squares). Error bars represent twice the standard error of the mean of one hundred parametric bootstrap replicates of the input data (site frequency spectra). All immune-control comparisons are significantly different from zero and each immune class is significantly different from its controls.

4.3.2 Population-level comparisons: the ancestral and three derived populations

A. Within-population analyses: characterizing genetic diversity and signatures of selection

We measured the genetic diversity of immune genes using both pairwise nucleotide diversity (π) and Watterson's θ within each of the four populations (Tables 4.1-4.4 and Fig. 4.6). Consistently across all four populations, the full set of immune genes did not show any significant differences compared to control genes at either 0-fold or 4-fold sites. For recognition genes, there was an overall trend toward higher genetic variation than controls at the 0-fold sites across populations; however, this was not statistically significant for any population. For signaling genes, there was an

overall consistent trend toward lower genetic variation than controls at both the 0-fold and 4-fold sites across populations. Notably, in all populations, both π and Watterson's θ were significantly lower than controls at the 0-fold sites of signaling genes. For modulation genes there was an overall trend toward higher genetic variation than controls across populations for both the 0-fold and 4-fold sites; however, this was not statistically significant for any of the four populations. For effector genes, the pattern was more variable, and no significant differences to the controls were found in any of the populations.

We measured signatures of selection within each population using Tajima's D (Tables 4.1-4.4 and Fig. 4.6). As a group, immune genes were not under uniformly strong directional or balancing selection in any population, with one exception: in the Atlantic population, the 0-fold sites exhibited significantly lower Tajima's D compared to the control genes, suggesting that, as a group, they experience increased directional selection. When considering genes of each functional class separately, there were differences in patterns not only between functional classes but also across populations. For recognition genes, the North America population showed a significantly lower Tajima's D at the 0-fold sites than controls, but this was not found in any other population (Florida was marginally significant). For signaling genes, the Atlantic population showed a significantly lower Tajima's D than controls at the 0-fold sites (Florida was marginally significant), but not at the 4-fold sites; in North America, Tajima's D was significantly lower than controls at the 4-fold sites, but not at the 0-fold sites. For modulation genes, no significant differences to the controls were found across populations. For effector genes, both the North America and Florida populations displayed significantly higher Tajima's D values compared to their controls: in North America, Tajima's D was significantly higher than controls at the 4-fold sites, while in Florida the 0-fold sites showed higher Tajima's D than controls.

Taken together, across all populations, immune genes as a group did not consistently exhibit significantly different levels of genetic variation and signatures of selection. Regarding genetic variation, a highly consistent pattern across populations was that the 0-fold sites of signaling genes showed significantly lower variation compared to control genes. There was also a trend for recognition and modulation genes to have greater variation than their respective controls. Regarding signatures of selection, the four populations exhibited moderately different patterns – there was no universal pattern across all populations. While effector genes displayed significantly higher Tajima's D than controls, indicating balancing selection in some populations, recognition and signaling genes showed significantly lower Tajima's D than their controls in some populations, indicating directional selection. Analyses based on all sites within each gene showed similar qualitative results (see supplemental information Tables S3-6 and Figs. S1-4).

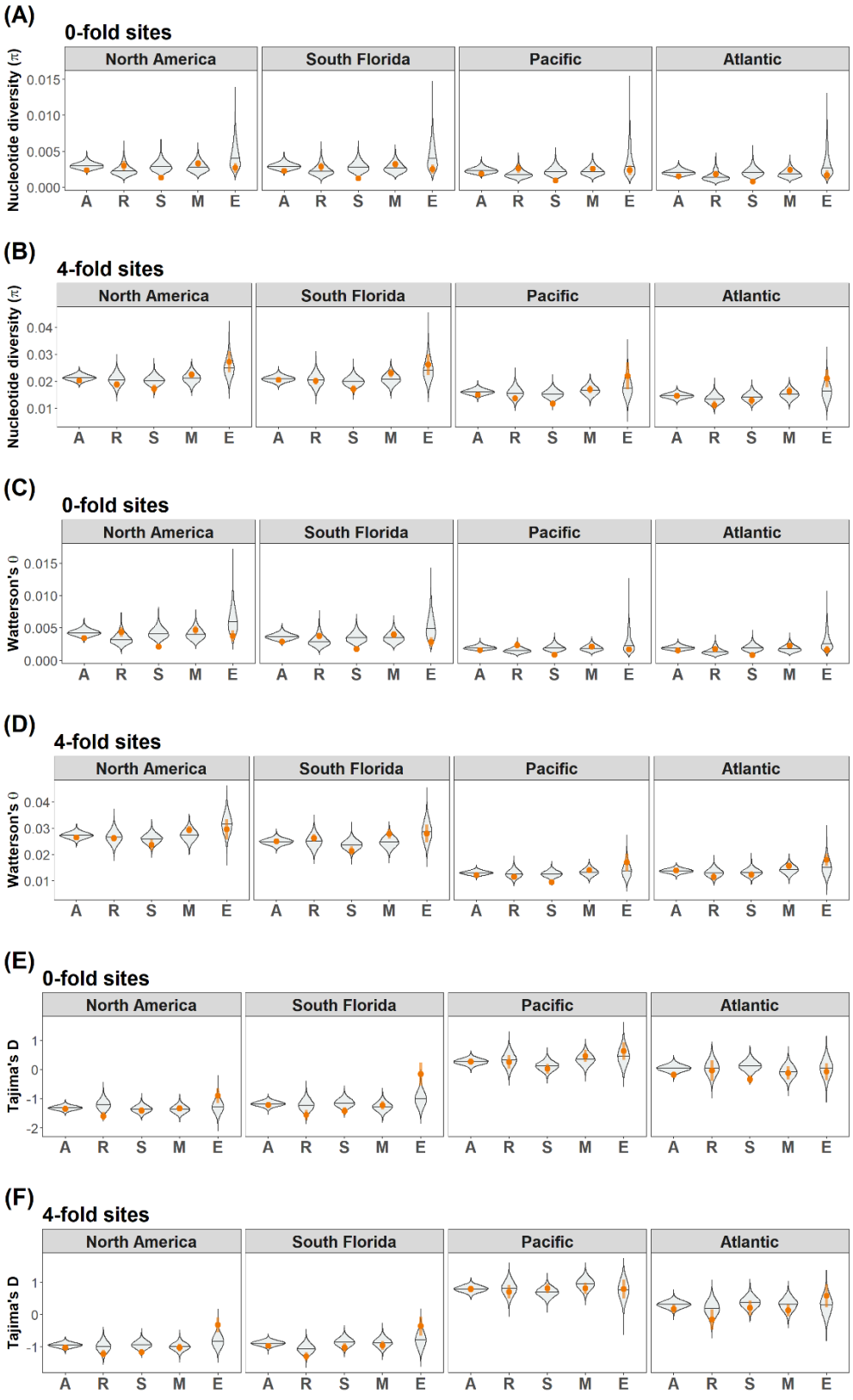

(A)
0-fold sites
North America
South Florida
Pacific
Atlantic
Nucleotide diversity (π)
0.015
0.010
0.005
0.000
A R S M E
(B)
4-fold sites
North America
South Florida
Pacific
Atlantic
Nucleotide diversity (π)
0.04
0.03
0.02
0.01
A R S M E
(C)
0-fold sites
North America
South Florida
Pacific
Atlantic
Watterson's θ
0.015
0.010
0.005
0.000
A R S M E
(D)
4-fold sites
North America
South Florida
Pacific
Atlantic
Watterson's θ
0.04
0.03
0.02
0.01
A R S M E
(E)
0-fold sites
North America
South Florida
Pacific
Atlantic
Tajima's D
1
0
-1
-2
A R S M E
(F)
4-fold sites
North America
South Florida
Pacific
Atlantic
Tajima's D
1
0
-1
A R S M E

Figure 4.6. Population genetic statistics of immune genes in all four populations (North America, Florida, Pacific, and Atlantic) using the paired-control approach. 0-fold sites shown in (A), (C), and (E), and 4-fold sites shown in (B), (D), and (F). (A) and (B): Nucleotide diversity ($\boldsymbol{\pi}$); (C) and (D): Watterson's θ; (E) and (F): Tajima's D. Each immune gene group was compared to selected pair-control sets. Violin plots show the distribution of the mean of each control set generated with 10, 000 permutations. The orange dots and vertical lines indicate mean ±1 SEM of the immune gene group of interest. X-axis represents immune gene groups: all immune genes (A), recognition genes (R), signaling genes (S), modulation genes (M), and effector genes (E).

Table 4.2. Population genetic statistics of immune genes in the Florida population using the paired-control approach. The upper half shows results based on the 0-fold sites and the lower half shows results based on the 4-fold sites. F_{ST} was compared to the North American population. "All immune" indicates the full immune gene set. In each statistic, the first row shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % are labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 are labeled in bold. Asterisks indicate: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
0-fold sites					
π: test statistic	-0.07	0.01	-0.07	0.01	-0.03
π: > 0 (%)	4.54	78.83	**0.05**	76.22	13.21
π: *P*-value.	0.118	0.461	**0.025***	0.498	0.291
Watterson's θ: test statistic	-0.08	0.02	-0.08	0.01	-0.03
Watterson's θ: > 0 (%)	3.44	82.81	**0.07**	73.52	5.98
Watterson's θ: *P*-value	0.090	0.346	**0.021***	0.558	0.161
Tajima's D: test statistic	-5.02	-6.80	-11.83	1.14	12.46
Tajima's D: > 0 (%)	29.28	3.40	2.78	60.66	**99.91**
Tajima's D: *P*-value	0.594	0.077	0.062	0.794	**0.001****
F_{ST}: test statistic	0.59	0.53	0.14	0.03	-0.11
F_{ST}: > 0 (%)	94.37	**99.84**	71.61	61.59	18.02
F_{ST}: *P*-value	0.098	**0.002****	0.575	0.837	0.349
4-fold sites					
π: test statistic	-0.05	-0.01	-0.13	0.06	0.02
π: > 0 (%)	34.57	44.02	3.79	85.30	66.74
π: *P*-value.	0.683	0.868	0.090	0.294	0.706
Watterson's θ: test statistic	0.00	0.02	-0.10	0.09	-0.02
Watterson's θ: > 0 (%)	50.89	69.13	10.34	93.40	38.79
Watterson's θ: *P*-value	0.979	0.629	0.209	0.129	0.764
Tajima's D: test statistic	-9.01	-4.80	-7.70	-2.76	6.26
Tajima's D: > 0 (%)	14.05	5.21	7.91	25.61	95.51
Tajima's D: *P*-value	0.281	0.107	0.165	0.501	0.083
F_{ST}: test statistic	-0.22	-0.12	0.05	-0.14	-0.02
F_{ST}: > 0 (%)	24.36	15.97	61.89	15.73	45.88
F_{ST}: *P*-value	0.470	0.321	0.810	0.314	0.849

Table 4.3. Population genetic statistics of immune genes in the Pacific population using the paired-control approach. The upper half shows results based on the 0-fold sites and the lower half shows results based on the 4-fold sites. F_{ST} was compared to the North American population. "All immune" indicates the full immune gene set. In each statistic, the first row shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % are labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 are labeled in bold. Asterisks indicate: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
0-fold sites					
π: test statistic	-0.05	0.02	-0.06	0.01	-0.02
π: > 0 (%)	11.28	90.28	**0.09**	72.45	32.43
π: *P*-value.	0.239	0.139	**0.032***	0.598	0.546
Watterson's θ: test statistic	-0.04	0.02	-0.05	0.01	-0.01
Watterson's θ: > 0 (%)	9.07	96.35	**0.04**	71.41	24.07
Watterson's θ: *P*-value	0.205	0.039	**0.022***	0.614	0.433
Tajima's D: test statistic	0.56	-1.27	-4.04	3.50	2.38
Tajima's D: > 0 (%)	51.93	39.59	27.39	73.21	69.72
Tajima's D: *P*-value	0.963	0.798	0.545	0.539	0.604
F_{ST}: test statistic	0.88	0.67	0.75	-0.19	-0.35
F_{ST}: > 0 (%)	77.94	88.78	87.07	39.64	26.83
F_{ST}: *P*-value	0.457	0.216	0.255	0.758	0.513
4-fold sites					
π: test statistic	-0.13	-0.04	-0.16	0.01	0.06
π: > 0 (%)	11.68	16.31	**0.55**	59.01	86.68
π: *P*-value.	0.239	0.320	**0.016***	0.844	0.251
Watterson's θ: test statistic	-0.10	-0.02	-0.14	0.02	0.04
Watterson's θ: > 0 (%)	10.54	23.08	**0.16**	68.92	84.08
Watterson's θ: *P*-value	0.218	0.443	**0.006****	0.635	0.313
Tajima's D: test statistic	-1.13	-2.10	4.99	-4.25	0.22
Tajima's D: > 0 (%)	46.02	31.94	78.49	21.43	51.46
Tajima's D: *P*-value	0.916	0.633	0.427	0.423	0.960
F_{ST}: test statistic	-1.42	-0.58	-0.32	-0.33	-0.19
F_{ST}: > 0 (%)	9.24	6.90	31.19	31.71	37.14
F_{ST}: *P*-value	0.195	0.158	0.599	0.606	0.694

Table 4.4. Population genetic statistics of immune genes in the Atlantic population using the paired-control approach. The upper half shows results based on the 0-fold sites and the lower half shows results based on the 4-fold sites. F_{ST} was compared to the North American population. "All immune" indicates the full immune gene set. In each statistic, the first row shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % are labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 are labeled in bold. Asterisks indicate: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
0-fold sites					
π: test statistic	-0.06	0.01	-0.06	0.02	-0.02
π: > 0 (%)	5.51	77.01	**0.03**	82.50	11.93
π: *P*-value.	0.142	0.567	**0.032***	0.346	0.280
Watterson's θ: test statistic	-0.05	0.01	-0.05	0.01	-0.02
Watterson's θ: > 0 (%)	6.29	77.86	**0.09**	83.54	10.69
Watterson's θ: *P*-value	0.151	0.517	**0.026***	0.327	0.262
Tajima's D: test statistic	-25.41	-1.59	-20.34	-1.48	-2.00
Tajima's D: > 0 (%)	**1.89**	39.22	**0.34**	41.69	32.56
Tajima's D: *P*-value	**0.040***	0.775	**0.008****	0.823	0.658
F_{ST}: test statistic	1.14	0.46	1.03	0.71	-1.06
F_{ST}: > 0 (%)	80.35	79.38	90.02	84.53	2.68
F_{ST}: *P*-value	0.403	0.427	0.192	0.303	0.075
4-fold sites					
π: test statistic	-0.01	-0.04	-0.06	0.03	0.06
π: > 0 (%)	47.29	9.14	20.01	73.74	89.11
π: *P*-value.	0.922	0.191	0.389	0.536	0.206
Watterson's θ: test statistic	0.01	-0.03	-0.04	0.04	0.03
Watterson's θ: > 0 (%)	53.52	15.93	26.19	81.28	78.35
Watterson's θ: *P*-value	0.950	0.315	0.509	0.377	0.446
Tajima's D: test statistic	-16.20	-6.92	-7.33	-5.96	4.00
Tajima's D: > 0 (%)	8.46	8.13	17.00	15.96	81.88
Tajima's D: *P*-value	0.164	0.156	0.337	0.314	0.369
F_{ST}: test statistic	0.65	0.16	1.26	-0.41	-0.35
F_{ST}: > 0 (%)	71.55	64.06	95.15	24.18	24.04
F_{ST}: *P*-value	0.573	0.739	0.084	0.466	0.459

B. Across-population analyses: population-level differentiations

We analyzed population differentiation using the ancestral population (*i.e.*, North America) as the reference population (Tables 1-4 and Fig. 7). The full set of immune genes used in this study did not display any significant differentiation compared to control genes. Across each functional class, there were no universal differences. However, there was an overall non-significant trend across populations at 0-fold sites: recognition genes showed higher F_{ST} than controls while effectors displayed lower F_{ST} than controls. Between the Florida and the ancestral populations, recognition genes showed significantly greater F_{ST} than controls; between the Atlantic and the ancestral populations, effector genes showed marginally significantly lower F_{ST} than controls. Analyses based on all sites within each gene showed similar qualitative results (see supplemental information Tables S3-6 and Figs. S1-4).

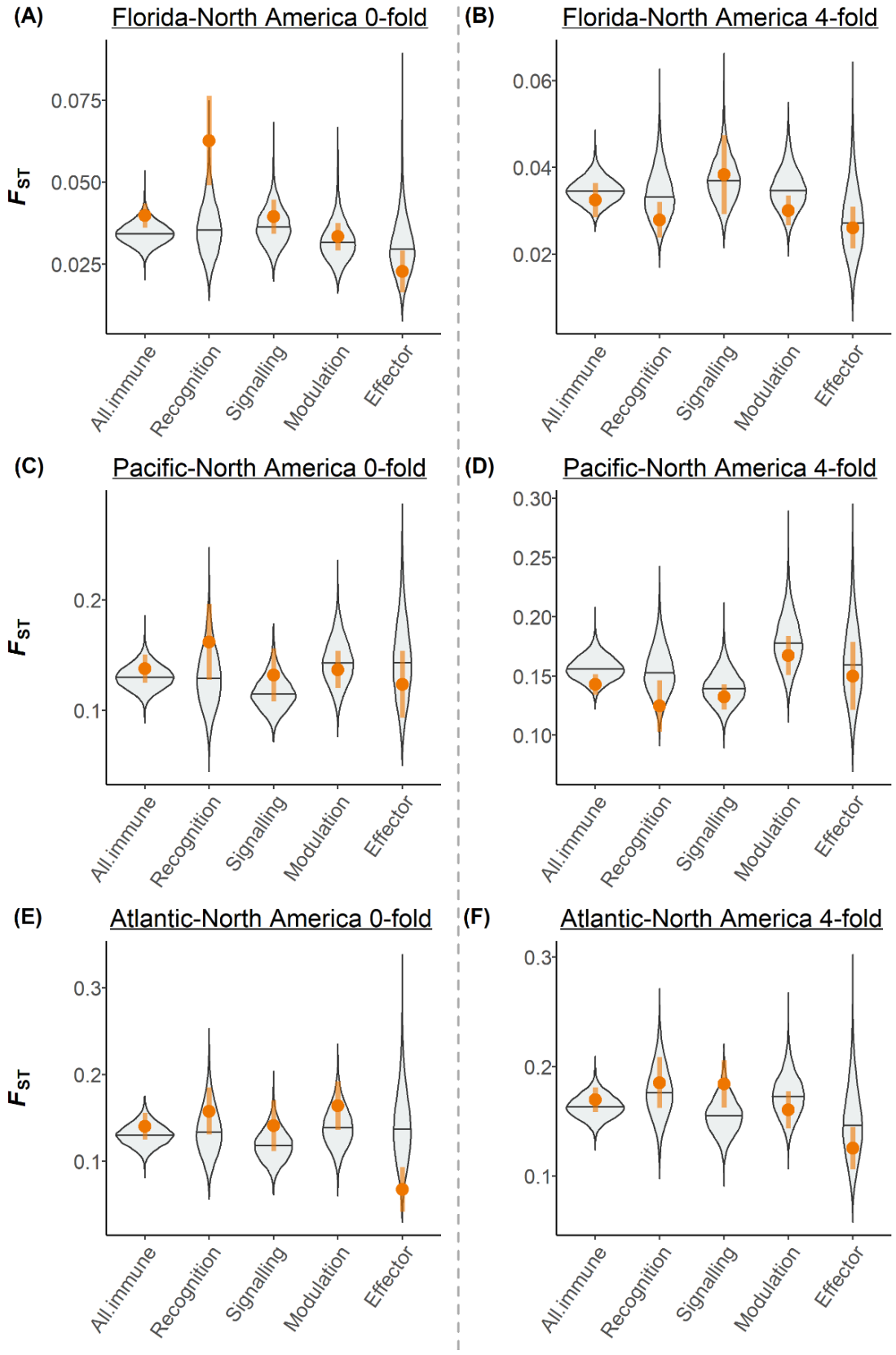

(A)
Florida-North America 0-fold
(B)
Florida-North America 4-fold
(C)
Pacific-North America 0-fold
(D)
Pacific-North America 4-fold
(E)
Atlantic-North America 0-fold
(F)
Atlantic-North America 4-fold
F_{ST}
All.immune
Recognition
Signalling
Modulation
Effector

Figure 4.7. F_{ST} of immune genes in each derived population compared to the ancestral (North American) population using the paired-control approach. 0-fold sites shown in (A), (C), and (E), and 4-fold sites shown in (B), (D), and (F). (A)-(B): South Florida population ($\boldsymbol{\pi}$); (C)-(D): Pacific population; (E)-(F): Atlantic population. Each immune gene group was compared to selected pair-control sets. Violin plots show the distribution of the mean of each control set generated with 10, 000 permutations. The orange dots and vertical lines indicate mean ±1 SEM of the immune gene group of interest.

4.3.3 Outlier analyses: access the patterns of outlier immune genes

We visualized Tajima's D and F_{ST} results together to assess outlier patterns, considering both signatures of selection and differentiation among populations simultaneously. In Fig. 4.8, outliers that fall into different areas suggest different evolutionary scenarios. Different immune gene functional groups did not seem to show distinct differences in outlier patterns, but they differed greatly in the proportion of genes that were outliers, ranging from 14.3% to 31.6% at 0-fold sites and from 7.1% to 42.9% at 4-fold sites. We separated outlier genes into five categories based on their location in the 2D Tajima's D- F_{ST} plot. We first compared whether the frequencies of outliers in each category (four outlier areas plus the central non-outlier area) differed across populations within each functional class. For the four functional classes, those frequencies did not differ significantly across populations at either the 0-fold or 4-fold sites (Table 4.5). Next, we compared whether the frequencies of outliers in each category differed across functional classes within each population. For the four populations, those frequencies did not differ significantly across functional classes at either the 0-fold or 4-fold sites (Table 4.6). In addition, we tested if immune genes, as one group, were disproportionally represented in genome-wide outliers, and found that they were not (see supplemental information Table S4.7). Overall, our results indicate no statistically significant differences in outlier patterns across populations or functional classes.

Table 4.5. Contingency tables of Tajima's D – F_{ST} outliers for each immune functional class. Numbers are counts of genes for each category. Q1 – Q4 categories represent the four quadrants shown in each Figure 8 plot (Q1 = top-right, Q2 = top-left, Q3 = bottom-left, Q4 = bottom-right). NS category represents non-outliers (i.e., the area within dotted gray lines). *P*-values from Fisher exact tests for each contingency table are shown in the last column. The North American population was not used because it was the reference group and did not have F_{ST} data.

Gene class	Sites	Population	Q1	Q2	Q3	Q4	NS	*P*-value
Recognition	0-fold	Florida	1	2	0	0	16	0.550
		Pacific	2	0	0	0	17	
		Atlantic	0	1	0	0	18	
Signaling	0-fold	Florida	2	2	1	0	36	0.923
		Pacific	1	1	2	0	37	
		Atlantic	0	2	2	0	37	
Modulation	0-fold	Florida	0	2	1	0	25	1.000
		Pacific	1	1	0	0	26	
		Atlantic	1	1	1	0	25	
Effector	0-fold	Florida	0	0	0	2	12	0.761
		Pacific	0	0	0	0	14	
		Atlantic	0	0	0	1	13	
Recognition	4-fold	Florida	0	0	0	0	19	0.766
		Pacific	0	0	1	0	18	
		Atlantic	1	1	0	0	17	
Signaling	4-fold	Florida	0	2	4	1	34	0.182
		Pacific	1	1	0	1	38	
		Atlantic	3	3	1	0	34	
Modulation	4-fold	Florida	0	0	0	0	28	1.000
		Pacific	0	0	0	0	28	
		Atlantic	0	1	0	0	27	
Effector	4-fold	Florida	0	0	0	2	12	0.365
		Pacific	0	0	0	0	14	
		Atlantic	0	1	1	1	11	

Table 4.6. Contingency tables of Tajima's D – F_{ST} outliers for each population. Numbers are counts of genes for each category. Q1 – Q4 categories represent the four quadrants shown in each Figure 8 plot (Q1 = top-right, Q2 = top-left, Q3 = bottom-left, Q4 = bottom-right). NS category represents non-outliers (i.e., the area within dotted gray lines). *P*-values from Fisher exact tests for each contingency table are shown in the last column. In the North American population, the analyses were based on only the Tajima's D data. Q1 represents "right area" and Q2 represents "left area". Q3 and Q4 were thus non-applicable.

Population	Sites	Gene class	Q1	Q2	Q3	Q4	NS	*P*-value
North America	0-fold	Recognition	0	0	NA	NA	19	1.000
		Signaling	0	2	NA	NA	39	
		Modulation	0	1	NA	NA	27	
		Effector	0	0	NA	NA	14	
Florida	0-fold	Recognition	1	2	0	0	16	0.374
		Signaling	2	2	1	0	36	
		Modulation	0	2	1	0	25	
		Effector	0	0	0	2	12	
Pacific	0-fold	Recognition	2	0	0	0	17	0.820
		Signaling	1	1	2	0	37	
		Modulation	1	1	0	0	26	
		Effector	0	0	0	0	14	
Atlantic	0-fold	Recognition	0	1	0	0	18	0.819
		Signaling	0	2	2	0	37	
		Modulation	1	1	1	0	25	
		Effector	0	0	0	1	13	
North America	4-fold	Recognition	0	0	NA	NA	19	0.105
		Signaling	0	2	NA	NA	39	
		Modulation	0	1	NA	NA	27	
		Effector	2	0	NA	NA	12	
Florida	4-fold	Recognition	0	0	0	0	19	0.092
		Signaling	0	2	4	1	34	
		Modulation	0	0	0	0	28	
		Effector	0	0	0	2	12	
Pacific	4-fold	Recognition	0	0	1	0	18	0.833
		Signaling	1	1	0	1	38	
		Modulation	0	0	0	0	28	
		Effector	0	0	0	0	14	
Atlantic	4-fold	Recognition	1	1	0	0	17	0.489
		Signaling	3	3	1	0	34	
		Modulation	0	1	0	0	27	
		Effector	0	1	1	1	11	

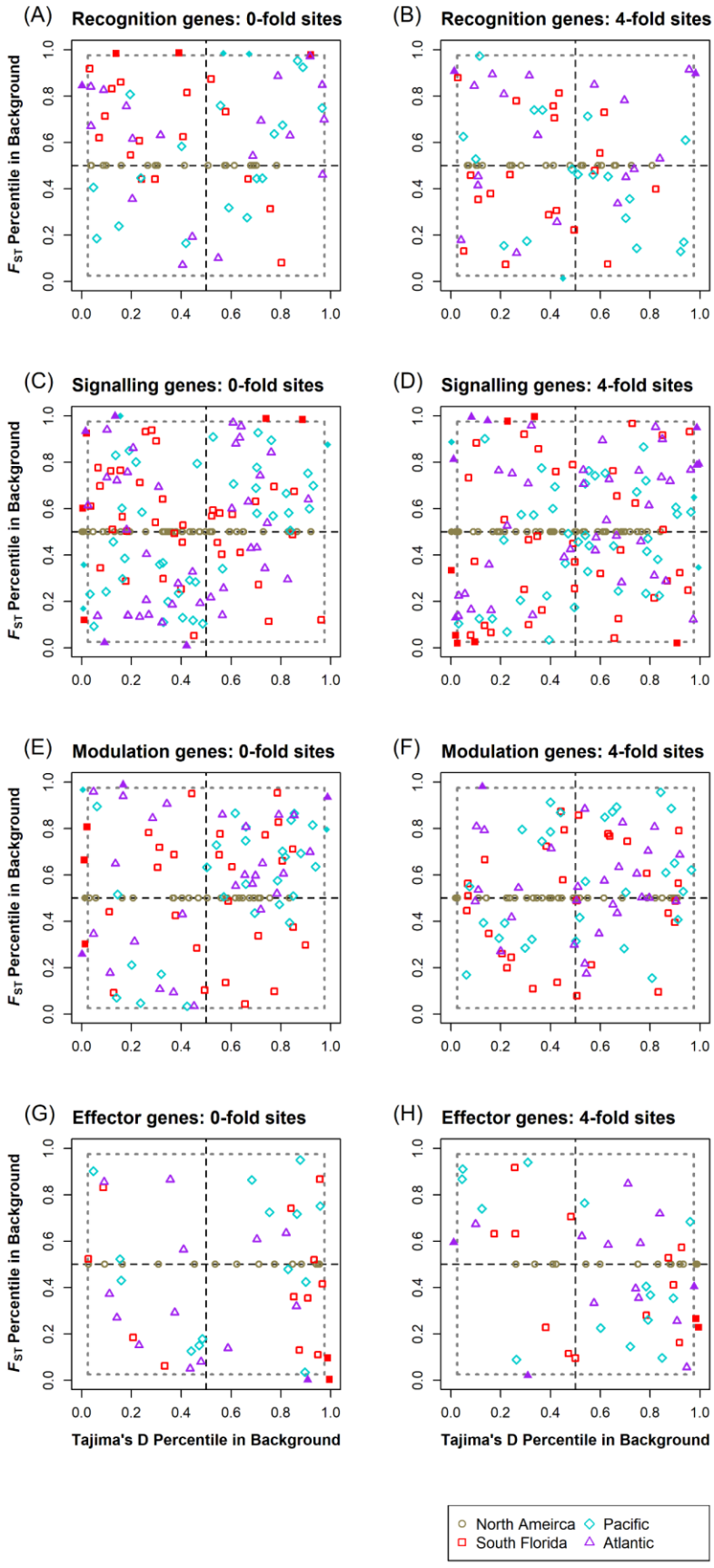

(A) Recognition genes: 0-fold sites
(B) Recognition genes: 4-fold sites
(C) Signalling genes: 0-fold sites
(D) Signalling genes: 4-fold sites
(E) Modulation genes: 0-fold sites
(F) Modulation genes: 4-fold sites
(G) Effector genes: 0-fold sites
(H) Effector genes: 4-fold sites
F_{ST} Percentile in Background
Tajima's D Percentile in Background
0.0 0.2 0.4 0.6 0.8 1.0
North Ameirca
South Florida
Pacific
Atlantic

Figure 4.8. Tajima's D – F_{ST} plots of the four immune gene functional classes. 0-fold sites shown in (A) – (D) and 4-fold sites shown in (E) – (H). (A) and (E): recognition (N = 19; 57.9% outlier in 0-fold; 31.6% outlier in 4-fold); (B) and (F): signaling (N = 41; 36.6% outlier in 0-fold; 53.7% outlier in 4-fold); (C) and (G): modulation (N = 28; 46.6% outlier in 0-fold; 17.9% outlier in 4-fold); (D) and (H): effector (N = 14; 64.3% outlier in 0-fold; 57.1% outlier in 4-fold). In each plot, populations were labeled in different colors and shapes. One dot represents one immune gene in one population, shown as their percentile in the genome background. Solid dots are outliers. Outliers were defined as < 2.5^{th} percentile or > 97.5^{th} percentile of the genome background. Dotted black lines indicate the median of genome background in each of the two measures, dividing the plot into four quadrants. Dotted gray lines indicate the boundaries of outlier areas. All data from the North American population, the reference population for F_{ST}, were plotted on the y = 0.5 horizontal line since they do not have F_{ST} results.

We identified individual immune genes that were genome outliers based on 0-fold sites and summarized their statistics across populations (Tajima's D outliers in Table 4.7, F_{ST} outliers in Table 4.8). Some genes exhibited distinct patterns across populations, as indicated by being outliers at different ends of the statistics. For example, *Pellino*, which belongs to the Toll pathway, was under directional selection (low Tajima's D) in the Florida population, while under balancing selection (high Tajima's D) in the Pacific population. One CLIP serine protease was under directional selection (low Tajima's D) in the Pacific population, while under balancing selection (high Tajima's D) in the Atlantic population. In addition, some of the patterns observed for F_{ST} outliers were population-specific – only shown in one population but not the others. Two out of three *Nimrod* genes were identified as F_{ST} outliers in the Pacific population compared to the ancestral population, and all of them showed higher differentiation (high F_{ST}). Two out of seven Scavenger receptor (SCR) genes were identified as F_{ST} outliers in the Florida population, and all of them showed higher differentiation (high F_{ST}). In contrast, some genes were identified as outliers in half of the populations in the same direction. For instance, one Toll-like receptor and *DOMELESS* were under directional selection (low Tajima's D) in both the North America and Florida populations at the 0-fold sites; one *Attacin*-like gene showed lower differentiation (low

F_{ST}) in the Florida and Atlantic population. However, no immune genes were consistently identified as outliers across all populations based on either Tajima's D or F_{ST}. Three genes were identified as outliers based on both Tajima's D and F_{ST}: Myeloid differentiation primary response 88 (*MyD88*), Protein inhibitor of activated STAT (*PIAS*), and one Attacin-like gene. *PIAS* showed a general trend of lower Tajima's D and lower F_{ST}, suggesting that it might be evolutionarily constrained. *MyD88* showed a general trend of higher F_{ST} and was an outlier in the Florida population. Also, in the Atlantic population, *MyD88* was a Tajima's D outlier, indicating directional selection. The Attacin-like gene showed a general trend of higher Tajima's D and was an outlier in the Florida population, indicating balancing selection. Also, it was an F_{ST} outlier in the Atlantic and Florida populations, indicating low differentiation. In summary, although our results did not reveal clear patterns of outliers based on functional groups or populations, individual outlier genes were identified. These results suggest that immune genes undergo individual evolutionary trajectories, and these trajectories vary across populations.

The analysis of outliers supports our notion that the complex evolutionary pressures have resulted in different patterns of selection on individual genes in the different populations, involving a wide variety of biological processes and targets. A few genes that showed high population differentiation (F_{ST} outliers at the upper end) are involved in cellular immune processes, such as phagocytosis. Two scavenger receptor genes showed high differentiation only in the Florida population, while two *nimrod* genes showed high differentiation only in the Pacific population. Notably, the *Nimrod* gene family is involved in recognizing foreign object for phagocytosis, which likely has direct interactions with pathogens (Estévez-Lao & Hillyer, 2014; Kurucz et al., 2007; Somogyi, Sipos, Pénzes, & Andó, 2010). Several of the outlier genes either belong to or interact with the Toll signaling pathway. For instance, two outlier genes encode Beta-1,3-glucan

recognition proteins (BGRPs), both of which recognize bacterial and/or fungal signals and are known to activate the toll signaling cascade in *Drosophila* (Kim et al., 2000). One of them is involved in activation of the phenoloxidase cascade (Matskevich, Quintin, & Ferrandon, 2010), while the other one leads to signal transmission that induces the expression of AMPs such as cecropin and attacin (Kim et al., 2000). Some members of the Toll signaling pathway, such as *spaetzle*, *Pellino*, and *MyD88*, were identified as outliers. *MyD88*, which is involved in regulating AMPs in *Drosophila* (Tauszig-Delamasure, Bilak, Capovilla, Hoffmann, & Imler, 2002). Attacins, which are AMPs against Gram-negative bacteria, are regulated mostly by the IMD pathway but also known to have some interactions with the Toll pathway (Tanji, Hu, Weber, & Ip, 2007).

Table 4.7. Summary of immune genes that are outliers according to Tajima's D at the 0-fold sites. Outliers were defined as < 2.5th percentile or > 97.5th percentile of the genome background. A gene was reported as an outlier when it met the criteria in at least one of the populations. The Tajima's D value of the 0-fold sites and the 4-fold sites of each outlier gene are shown. Values that are less than or equal to the genome median (i.e., 50th percentile) are underscored; values that are greater than the genome median are in italics. Values that are outliers are in bold. Genes that are reported as outliers in both Tajima's D and F_{ST} (Table 8) are colored in red. A Tajima's D value close to 0 indicates neutrality. A more negative Tajima's D value represents an excess of low-frequency polymorphisms than expectation, which indicates directional selection or population expansion. A more positive Tajima's D value represents low levels of both low- and high-frequency polymorphisms, which indicates balancing selection or population contraction.

Gene name	Gene number	Functional class	North America		Florida		Pacific		Atlantic	
			0-fold	4-fold	0-fold	4-fold	0-fold	4-fold	0-fold	4-fold
BGRP-like	DPOGS212941	Recognition	-1.74	-1.93	-1.90	-2.13	*0.59*	0.05	**-2.66**	-0.72
CLIP-like	DPOGS204835	Modulation	-2.42	**-2.28**	-1.79	-1.72	*0.07*	0.40	**-2.72**	*0.06*
CLIP-like	DPOGS204146	Modulation	*-1.13*	-1.17	*-0.37*	-1.30	**-2.08**	-0.14	***2.67***	-0.20
CLIP-like	DPOGS208169	Modulation	**-2.54**	*-0.91*	-2.21	*-0.67*	-0.67	-0.19	-0.26	*1.90*
CLIP-like	DPOGS211355	Modulation	-2.01	-1.81	**-2.67**	-1.56	-0.53	-0.03	-0.32	*0.74*
CLIP-like	DPOGS214570	Modulation	-1.43	*-0.79*	*-0.21*	*0.26*	***2.44***	*0.72*	-0.69	-0.78
CLIP-like	DPOGS206224	Modulation	-2.30	-1.31	**-2.64**	-1.78	*-0.06*	-0.39	-0.98	-1.49
CLIP-like	DPOGS205206	Modulation	*-0.81*	-1.28	**-2.59**	-1.62	-0.92	*1.17*	-1.25	*0.04*
Toll-like receptors	DPOGS205295	Signaling - Toll	**-2.60**	*-0.80*	**-2.77**	*0.31*	-0.97	*0.88*	*0.33*	*1.19*
MyD88	DPOGS205936	Signaling - Toll	*-0.29*	*-0.61*	*-0.07*	*-0.43*	*0.99*	***2.50***	**-2.21**	**-2.45**
Pellino	DPOGS214647	Signaling - Toll	*-0.95*	-1.88	**-2.60**	-1.37	***2.54***	*0.83*	-0.46	-1.09
Hem	DPOGS208954	Signaling - JNK	-1.42	*-0.80*	-1.54	-1.34	**-1.98**	0.70	-0.62	-0.06
PIAS	DPOGS214325	Signaling - JAK-STAT	-1.58	-1.09	-1.40	-1.37	**-1.93**	*0.87*	-0.30	*1.57*
DOMELESS	DPOGS200349	Signaling - JAK-STAT	**-2.79**	*-0.18*	**-2.68**	*-0.08*	-1.24	*1.70*	-0.83	*0.73*
Attacin-Like	DPOGS213997	Effector	*0.23*	*-0.22*	***2.03***	-1.00	*2.06*	*0.82*	*1.83*	*0.52*
PPO-like	DPOGS200017	Effector	*-0.74*	-1.30	***1.62***	-1.19	-1.48	-0.42	-0.38	**-2.43**

Table 4.8. Summary of immune genes that are outliers according to F_{ST} at the 0-fold sites. Outliers were defined as < 2.5th percentile or > 97.5th percentile of the genome background. A gene was reported as an outlier when it met the criteria in at least one of the population pairs. The F_{ST} value of the 0-fold sites and the 4-fold sites of each outlier gene are shown. Values that are less than or equal to the genome median (i.e., 50th percentile) are underscored; values that are greater than the genome median are in italics. Values that are outliers are in bold. Genes that are reported as outliers in both Tajima's D (Table 7) and F_{ST} are colored in red. F_{ST} is a measure of population differentiation due to genetic structure, with a value ranging from 0 to 1. An F_{ST} value equals to zero indicates no differentiation. An F_{ST} value equals to one indicates complete differentiation; different alleles are fixed in different populations.

Gene name	**Gene number**	**Functional class**	**Florida – North America**		**Pacific – North America**		**Atlantic – North America**	
			0-fold	4-fold	0-fold	4-fold	0-fold	4-fold
BGRP-like	DPOGS212940	Recognition	***0.14***	0.02	*0.20*	0.12	*0.14*	0.03
Class B-like SCR	DPOGS203180	Recognition	***0.12***	0.01	*0.17*	0.10	*0.15*	0.11
Other SCR	DPOGS214397	Recognition	***0.15***	*0.05*	0.07	*0.20*	*0.46*	*0.30*
NIM-like	DPOGS210210	Recognition	*0.06*	*0.04*	***0.48***	*0.20*	*0.28*	0.09
NIM-like	DPOGS210211	Recognition	*0.04*	*0.05*	***0.50***	*0.43*	*0.25*	*0.30*
CLIP-like	DPOGS215098	Modulation	*0.04*	*0.03*	*0.15*	*0.15*	***0.57***	***0.48***
SPZ-like	DPOGS209810	Signaling - Toll	0.02	**-0.01**	0.03	0.03	**0.00**	0.06
MyD88	DPOGS205936	Signaling - Toll	***0.14***	*0.10*	*0.27*	*0.17*	*0.35*	*0.24*
JNK	DPOGS213169	Signaling - JNK	***0.16***	***0.26***	0.06	*0.15*	0.02	0.03
PIAS	DPOGS214325	Signaling - JAK-STAT	0.00	0.01	0.05	*0.21*	**-0.01**	0.08
Stat	DPOGS212956	Signaling - JAK-STAT	*0.04*	*0.06*	***0.86***	*0.29*	***0.80***	***0.64***
Attacin-Like	DPOGS213997	Effector	**-0.02**	0.00	*0.17*	*0.21*	**-0.02**	*0.15*

4.4 Discussion

Even though immune system mediated defenses are well-studied in several insect species, population genomic analyses on the full set of insect canonical immune genes are rarely studied. In this study, we examined the evolutionary patterns of canonical immune genes within and among populations of monarch butterflies, and between monarch butterflies and a closely-related species. Our results demonstrate that immune genes, as one group, do not exhibit uniform patterns of selection, differentiation, or high genetic variation; different function classes show different patterns. Monarchs recently spread around the world via three main dispersal events (Pierce et al., 2014; Zhan et al., 2014)(Fig. 4.1). During these colonization processes across the Pacific, the Atlantic and into Central/South America, they have encountered different ecological conditions that are likely to drive the evolution of immune genes. Our results show that patterns of evolutionary change in immune genes of different functional groups vary to some extent across populations, suggesting that populations might not be under a uniform selection regime. This is further supported by assessing individual genes that are genome outliers, as some of them exhibit distinct differences across populations. Here, we compare our results to two other population genomic studies on the full set of immune genes – one on *Drosophila melanogaster* (Diptera) and the other on *Pieris napi* (Lepidoptera)(Early et al., 2017; Keehnen et al., 2018), discuss similarities and differences between patterns observed within species and between species comparisons, and discuss those evolutionary patterns in the context of ecological differences among those populations.

4.4.1 Population genomic patterns and adaptive evolution across different functional classes

Consistent with our results here, a limited body of work has demonstrated that different

components of the canonical insect immune system can face distinct selection pressures. Genes encoding proteins in the core signaling pathways, for example, have been shown to be more functionally constrained (Sackton et al., 2007). Consistent with this, low genetic variation in signaling genes is one of the most consistent patterns found in our study – signaling genes showed significantly lower genetic variation than control genes in all of the populations studied. Most likely, this reflects the increased removal of deleterious alleles among these loci. The DFE of signaling genes also points to this phenomenon, indicating a much larger proportion of strongly deleterious variants among new mutations relative to control genes. Broadly increased purifying selection can also help to explain the greater α value, which indicates the proportion of adaptive amino acid substitutions between species. If most new mutations are removed by purifying selection, then any divergence observed should primarily reflect adaptation, not neutral divergence (i.e. drift), even though the absolute amount of divergence might be quite small. Indeed, such increased purifying selection, if consistent over long periods of time, should reduce overall divergence between species. However, while signaling genes do have reduced average divergence relative to controls, this difference is not significant. Thus, it is possible that the strong purifying selection we observe in *D. plexippus* is a relatively recent phenomenon that manifests patterns in population diversity but not yet at the level of species divergence. Further population genetic analysis in other *Danaus* species would be required to assess whether there are long-term patterns of selection for this group of butterflies. Given the broad finding of functional constraint in other distantly related species, such variability in evolutionary pressures among signaling genes between closely related species is an intriguing possibility.

In striking contrast to signaling genes, modulation genes show a consistent pattern of increased diversity. While nucleotide diversity is only somewhat elevated, and not significantly so,

interspecific divergence is greatly increased. One good explanation for these patterns is that modulation genes experience relaxed selection compared to controls. This idea fits well with patterns in the DFE, which indicates notably more neutral variants and fewer strongly deleterious variants among new mutations among modulation genes. The relatively rapid divergence of modulation genes due to fixation of neutral or weakly deleterious mutations can explain the reduced α value. Taken together, signaling and modulation genes both exhibited consistent evolutionary patterns across populations, suggesting that the selection regime on these two functional classes might not differ strongly across populations.

Signaling genes and modulation genes are sometimes considered as one functional class (*e.g.*, Waterhouse et al., 2007), but our results show distinct differences in genetic diversity. Signaling genes, especially those within the Toll, IMD, JAK-STAT, and JNK pathways, are well-characterized for their function. However, relatively little is known about the functional roles of modulation genes, most of which are CLIP serine proteases (Lemaitre & Hoffmann, 2007). Our results suggest that signaling and modulation genes likely have different functional roles, as they exhibit notably distinct patterns of selection.

In contrast, genes that encode proteins that have direct interactions with pathogens, such as recognition and effector genes, have been shown to evolve more rapidly as they are more likely targets of host-pathogen coevolutionary arms races (Sackton et al., 2007). For effector genes, some recent studies have demonstrated signatures of balancing selection, especially for AMPs (Chapman et al., 2018; Unckless, Howick, & Lazzaro, 2016; Unckless & Lazzaro, 2016). Similarly, in monarchs, effector genes show notable evidence of balancing selection. Specifically, Tajima's D is elevated, at least in some populations. However, this occurs without a clear signal of broadly elevated heterozygosity, which would be expected in many scenarios involving balancing selection.

This might result if a few effector genes show strongly balanced patterns, contributing substantially to greater Tajima's D but less so to average variation across effector loci. Anticipating or interpreting the DFE under balancing selection is not straightforward (Connallon & Clark, 2015). Yet it is very clear that the DFE is qualitatively distinct between effectors and their controls, as well as the other classes of immune genes: there appear to be many fewer new mutations showing strongly deleterious effects. One plausible explanation for this is that balancing selection counteracts or retards the efficient removal of deleterious variants, producing a relative excess of intermediate frequency polymorphisms in SFS, manifesting as an abatement of purifying selection in this DFE analysis. Whatever the cause, these distinct population genetic patterns among effectors do not obviously carry over to patterns of divergence, where there is little difference from controls.

We observed considerable differences in patterns of selection across populations on effector genes. Specifically, signatures of balancing selection were observed in the North America and Florida populations but not in the Pacific and Atlantic populations. One possible interpretation is that this pattern reflects a shift in selective regime among populations. When monarchs dispersed to distant locations across the Pacific and Atlantic oceans, the selection regimes shifted toward either directional selection or were relaxed, leading to a loss of selective signal in these two populations. Alternatively, the selective regime may be constant, but demographic effects, including bottlenecks and other non-equilibrium effects, are masking the signal. Specifically, bottleneck effects, which the Pacific and Atlantic populations have experienced (Pierce et al., 2014; Zhan et al., 2014), can skew allele frequencies. The effect of skewed allele frequencies due to bottlenecks can reduce or mask the signal of balancing selection. In a more extreme scenario, one of the selected variants could be entirely removed by bottlenecks so that the balanced loci cannot

be restored after the population recovered. Even though we tried to account for demographic effects by using a paired-control approach, there is still a possibility that we have a reduced resolution in the derived populations due to demographic effects.

Evolutionary analyses of immune genes in other species, particularly *Drosophila*, indicate that recognition genes have the strongest evidence for adaptive evolution among immune functional groupings (McTaggart et al., 2012; Sackton et al., 2007). By comparison, there was distinctly mixed evidence for strong selection among recognition genes in monarchs. In north America (and Florida), Tajima's D was notably reduced relative to controls for both 0x and 4x sites, though without much reduction in 4x heterozygosity, and even a modest increase for 0x heterozygosity. If this pattern reflects recent selective sweeps among some recognition genes for these populations, it is likely a narrow range of parameters that would produce such skewed distributions of diversity (i.e. Tajima's D) without also affecting the amount of diversity (i.e. heterozygosity). Nonetheless, recurring adaptation among recognition genes could also explain the significantly elevated Dn/Ds observed in divergence to *D. gillipus*. Alternatively, this could result from relaxed constraint, as we argued above for modulators. Also, like modulators, the DFE of recognition genes suggests relatively fewer strongly selected variants compared to controls, and α is also lower. The mixed signals for selection in recognition loci also play out among patterns of population differentiation. 0x F_{ST} between north American and Florida populations is strongly elevated relative to controls; a similar but less extreme signal occurs for Pacific vs. North America. While this might be interpreted as evidence for local adaptation among these distinct populations, no such pattern was observed among linked 4x sites, which are expected to show the same pattern. These contrasting patterns among the different analytical components employed here are not easily synthesized into a single coherent biological interpretation for recognition loci; a more detailed,

gene-by-gene analysis may be required to resolve many of these discrepancies.

We also observed differences in patterns of selection across populations on recognition genes. Specifically, significant signatures of directional selection were observed in the North America population, but they were marginally significant in the Florida population, and not significant in the Pacific and Atlantic populations. Intriguingly, this pattern across populations is similar to what was observed for the balancing selection on effector genes. One possible interpretation is that this pattern reflects a shift in selective regime among populations. That is, the differences reflect local adaption to pathogens or parasites. Alternatively, the selective regime may be constant, but demographic effects are masking the signal. Bottleneck effects can exert similar effects as selective sweeps, removing rare alleles, but acting across the entire genome background instead. The removal of rare alleles can result in a disproportional loss of genetic variation on loci with high- and intermediate-level polymorphisms compared to loci under directional selection, which already have lower polymorphism. That is, bottlenecks can result in a disproportional loss of genome-wide genetic variation compared to loci under directional selection. Similar to directional selection, selection sweeps can result in a low Tajima's D value by removing rare alleles (Nielsen & Slatkin, 2013). Therefore, although we tried to account for demographic effects in our analyses, there is still a possibility that we have a reduced resolution in the derived populations.

Overall, our results and those of previous studies on *Drosophila melanogaster* and *Pieris napi* (Early et al., 2017; Keehnen et al., 2018) highlight that it may be common for different components of the insect canonical immune system to have different evolutionary trajectories. A common trend among the three taxa is that genes within signaling pathways show lower levels of genetic variation, genes involve in recognition show higher levels of population differentiation in some scenarios (Early et al., 2017; Keehnen et al., 2018), and that genes encoding effector molecules (especially

AMPs) show signatures of balancing selection (Chapman et al., 2018; Keehnen et al., 2018; Unckless et al., 2016). The emergence of these common patterns across insect species that differ considerably in life histories and taxonomy suggests that there may be some general evolutionary patterns among insect immune genes.

4.4.2 Ecological differences among populations and their potential consequences for immune gene evolution

Ecological factors that vary across populations affect the strength and type of selection and can therefore lead to local adaptation (Eizaguirre et al., 2012). When monarch butterflies dispersed around the world, they experienced novel ecological conditions, likely resulting in differential selection across populations. First, different populations face different pathogen and parasite pressures. In monarchs, the most common and best-understood parasite is the virulent specialist protozoan parasite *Ophryocystis elektroscirrha*, which occurs at low prevalence in the ancestral North American population but at much greater prevalence in the tropical and sub-tropical locations that monarchs colonized during their worldwide dispersal (Altizer & de Roode, 2015), resulting in greater parasitism risk and possibly stronger selection of monarch immunity. Second, although North American monarchs migrate thousands of kilometers to overwinter in Central Mexico, the derived populations that established during world-wide dispersal have become non-migratory (Zhan et al., 2014). This loss of migration is likely partly responsible for the increased parasite prevalence in derived populations. In North America, the strenuous annual migration weeds out heavily infected monarchs, thus reducing parasite prevalence. In non-migratory populations, this seasonal break on parasite transmission has been eliminated, leading to greater transmission and prevalence (Altizer & de Roode, 2015; Altizer, Hobson, Davis, De Roode, &

Wassenaar, 2015; Bartel, Oberhauser, de Roode, & Altizer, 2011). Although this greater prevalence may select for greater immunity, it is also possible that the lack of a migratory phase, and the accompanying lack of a generation that needs to survive for long periods of time as it flies thousands of kilometers, results in less investment in immunity. Third, while the majority of North American monarchs utilize *Asclepias syriaca* (common milkweed) as their larval host plant species, monarchs in newly colonized populations rely on other species, including *Asclepias curassavica*, *A. fruticosa*, and *A. physocarpa*. Notably, these species have greater concentrations of cardenolides (secondary toxic compounds), which have been shown to reduce *O. elektroscirrha* infection, growth and virulence (Gowler et al., 2015; Sternberg et al., 2012; Tao, Hoang, et al., 2016). The use of such medicinal compounds could in theory relax selection on immune genes, especially when immune responses are costly (de Roode et al., 2013; Evans et al., 2006; Gerardo et al., 2010; Parker et al., 2011). Finally, while we know most about parasitism by *Ophryocystis elektroscirrha,* monarchs are undoubtedly challenged by a suite of pathogens and parasites that vary in presence and prevalence across populations. These differences in disease pressure undoubtedly shape the evolution of monarch immune defenses.

The different ecological conditions experienced by monarchs as they dispersed around the world do not act in isolation, resulting in a complex mosaic of factors that simultaneously select for greater or lesser investment in immunity. Furthermore, the evolutionary patterns of immune gene evolution also may be influenced by demographic history and stochastic processes. In our analyses, immune genes as a group do not display consistent patterns across populations. For instance, directional selection on recognition genes, which indicates an excess of rare alleles, was only seen in the North America population (Florida was marginally significant). Furthermore, different immune genes were outliers in different populations. This difference among populations

could in part be driven by genetic drift rather than differential selection; however, few immune genes were identified as outliers in multiple populations with strikingly different patterns. For example, *Pellino*, which belongs to the Toll pathway, showed an excess of rare alleles in the Florida population (directional selection) but showed maintenance of multiple alleles at moderate frequency (balancing selection) in the Pacific population, indicating that the selection forces between these two populations are very different.

4.5 Conclusions

In summary, our results demonstrated that immune genes as a whole were not under uniform patterns of selection or differentiation compared to the genome background. Different components of the immune system exhibit different evolutionary patterns. Signaling genes exhibited consistently low levels of genetic variation across populations and between the two *Danaus* species, indicating they are likely very constrained, while modulation genes exhibited the opposite pattern - signatures of relaxed selection. In contrast, effector and recognition genes exhibited less consistent patterns across populations. In some populations, effector genes exhibited signatures of balancing selection, while recognition genes exhibited directional selection and population differentiation. We found some clear differences among populations on individual genes that are genomic outliers, suggesting that immune genes undergo individual evolutionary trajectories. To a lesser extent, we also found some population-specific differences when considering each functional class separately. These results support the hypothesis that monarch populations do not face uniform selection pressures on immune genes.

The identification of immune genes that are under differential selection in monarch populations opens the way for further functional and ecological characterization. Specifically,

population-specific patterns indicate a possibility of local adaptation, and functional characterization is needed to understand the phenotypic effects of different alleles of immune genes, especially as they relate to important ecological factors, such as the prevalence of *O. elektroscirrha* and the use of medicinal milkweeds. Such functional characterization is also needed because several insect canonical immune genes, especially signaling genes, have pleiotropic functions in immunological and non-immunological processes (Lemaitre & Hoffmann, 2007). Therefore, evolutionary patterns on those genes may not be solely driven by selection pressures on immunity.

4.6 Acknowledgments

We thank William J. Palmer for providing immune gene annotation resources for *Danaus plexippus*. We thank Robert L. Unckless, Joanna Chapman, Erika Diaz-Almeyda, Amanda A. Peirce, Yaw Kumi-Ansu, Venkat Talla, and the KU EEB-genetics group for helpful discussion on this research, and Robert A. Pettit III and Yu-Hui Lin for helpful guidance and discussion on bioinformatics computing. Analyses were performed on resources provided by the University of Kansas Information and Telecommunication Technology Center. This work was supported by National Science Foundation (NSF) grant IOS-1557724 to JCdR and NMG, and NSF grant DEB-1457758 to JRW.

4.7 Supplemental Information

Table S4.1. *Danaus* samples used in this study. Sample information obtained from Zhan *et al.* (2014). We assigned each *D. plexippus* samples into genetic populations based on their collecting location.

Species	Accession	Sample	Sex	Collecting location	Assigned population
D. plexippus	SRX681756	Plex_CA_W48_F	female	Pismo Beach, CA, USA	North America
D. plexippus	SRX680103	Plex_FLn_StM109_M	male	St. Marks, FL, USA	North America
D. plexippus	SRX680104	Plex_FLn_StM122_F	female	St. Marks, FL, USA	North America
D. plexippus	SRX680105	Plex_FLn_StM123_F	female	St. Marks, FL, USA	North America
D. plexippus	SRX681753	Plex_FLn_StM146_M	male	St. Marks, FL, USA	North America
D. plexippus	SRX679271	Plex_MA_HI012_F	female	Massachusetts, USA	North America
D. plexippus	SRX679303	Plex_MA_HI023_M	male	Massachusetts, USA	North America
D. plexippus	SRX679310	Plex_MA_HI035_F	female	Massachusetts, USA	North America
D. plexippus	SRX681737	Plex_MEX_1742_F	female	Cerro Pelon, Mexico	North America
D. plexippus	SRX679305	Plex_NJ_116_M	male	New Jersey, USA	North America
D. plexippus	SRX679306	Plex_NJ_203_F	female	New Jersey, USA	North America
D. plexippus	SRX680106	Plex_TX_T11_F	female	Texas, USA	North America
D. plexippus	SRX681744	Plex_FLs_MIA2514_M	male	Miami, FL, USA	South Florida
D. plexippus	SRX681537	Plex_FLs_MIA126_F	female	Miami, FL, USA	South Florida
D. plexippus	SRX681742	Plex_FLs_MIA16_M	male	Miami, FL, USA	South Florida
D. plexippus	SRX681757	Plex_FLs_MIA11_F	female	Miami, FL, USA	South Florida
D. plexippus	SRX681758	Plex_FLs_MIA122_M	male	Miami, FL, USA	South Florida
D. plexippus	SRX681746	Plex_FLs_MIA40_F	female	Miami, FL, USA	South Florida
D. plexippus	SRX681747	Plex_FLs_MIA454_F	female	Miami, FL, USA	South Florida
D. plexippus	SRX680118	Plex_WSM_M36_M	male	Samoa	Pacific
D. plexippus	SRX681528	Plex_WSM_M38_F	female	Samoa	Pacific
D. plexippus	SRX681647	Plex_FJI_M81_M	male	Fiji	Pacific
D. plexippus	SRX681648	Plex_FJI_M82_M	male	Fiji	Pacific
D. plexippus	SRX681541	Plex_NCL_M122_M	male	New Caledonia	Pacific
D. plexippus	SRX681543	Plex_NCL_M128_M	male	New Caledonia	Pacific
D. plexippus	SRX681765	Plex_AUS_18_F	female	Australia	Pacific
D. plexippus	SRX681766	Plex_AUS_87_M	male	Australia	Pacific
D. plexippus	SRX681532	Plex_AUS_99_F	female	Australia	Pacific
D. plexippus	SRX681767	Plex_NZL_16_M	male	New Zealand	Pacific
D. plexippus	SRX681538	Plex_NZL_3_F	female	New Zealand	Pacific
D. plexippus	SRX681768	Plex_NZL_4_F	female	New Zealand	Pacific
D. plexippus	SRX680109	Plex_ESP_28_M	male	Spain	Atlantic
D. plexippus	SRX680110	Plex_ESP_29_F	female	Spain	Atlantic
D. plexippus	SRX680111	Plex_ESP_30_M	male	Spain	Atlantic
D. plexippus	SRX681546	Plex_MAR_M09_F	female	Morocco	Atlantic
D. plexippus	SRX681545	Plex_MAR_M08_M	male	Morocco	Atlantic

Species	**Accession**	**Sample**	**Sex**	**Collecting location**	**Assigned population**
D. plexippus	SRX681544	Plex_MAR_M034_M	male	Morocco	Atlantic
D. gilippus	SRX998564	Gili_TX_01_F	female	Texas, USA	N/A

Table S4.2. List of immune genes used in this study. Note that immune genes on sex chromosomes were not included in the analyses.

Gene ID	Gene name	Gene length	Functional class	Chromosome
DPOGS200905	PGRP-like	3371	Recognition	Z-chromosome
DPOGS209813	PGRP-like	1574	Recognition	Autosome
DPOGS206909	PGRP-like	1472	Recognition	Z-chromosome
DPOGS206910	PGRP-like	7919	Recognition	Z-chromosome
DPOGS207148	PGRP-like	15320	Recognition	Z-chromosome
DPOGS209814	PGRP-like	2222	Recognition	Autosome
DPOGS206026	PGRP-like	3333	Recognition	Z-chromosome
DPOGS212963	BGRP-like	7393	Recognition	Autosome
DPOGS215599	BGRP-like	2910	Recognition	Autosome
DPOGS212940	BGRP-like	4002	Recognition	Autosome
DPOGS212941	BGRP-like	3544	Recognition	Autosome
DPOGS212964	BGRP-like	3508	Recognition	Autosome
DPOGS212965	BGRP-like	3969	Recognition	Autosome
DPOGS203317	Frep-like	5328	Recognition	Z-chromosome
DPOGS206045	Frep-like	4504	Recognition	Z-chromosome
DPOGS203951	Frep-like	10862	Recognition	Z-chromosome
DPOGS210549	Class B-like SCR	5148	Recognition	Autosome
DPOGS203180	Class B-like SCR	13541	Recognition	Autosome
DPOGS202796	Class B-like SCR	9956	Recognition	Autosome
DPOGS214397	Other SCR	13196	Recognition	Autosome
DPOGS213636	Other SCR	15159	Recognition	Autosome
DPOGS212634	Other SCR	14283	Recognition	Autosome
DPOGS202826	Other SCR	9352	Recognition	Autosome
DPOGS215836	TEP-like	10955	Recognition	Autosome
DPOGS210251	NIM-like	13838	Recognition	Autosome
DPOGS210210	NIM-like	5318	Recognition	Autosome
DPOGS210211	NIM-like	13481	Recognition	Autosome
DPOGS204835	CLIP-like	3824	Modulation	Autosome
DPOGS205231	CLIP-like	27491	Modulation	Autosome
DPOGS206561	CLIP-like	4655	Modulation	Autosome
DPOGS215180	CLIP-like	11215	Modulation	Autosome
DPOGS215181	CLIP-like	10878	Modulation	Autosome
DPOGS206562	CLIP-like	14594	Modulation	Autosome
DPOGS206563	CLIP-like	3009	Modulation	Autosome
DPOGS213841	CLIP-like	3609	Modulation	Autosome
DPOGS215183	CLIP-like	13397	Modulation	Autosome
DPOGS215188	CLIP-like	4339	Modulation	Autosome
DPOGS204146	CLIP-like	5061	Modulation	Autosome

Gene ID	Gene name	Gene length	Functional class	Chromosome
DPOGS204147	CLIP-like	6895	Modulation	Autosome
DPOGS201678	CLIP-like	3492	Modulation	Autosome
DPOGS215220	CLIP-like	3818	Modulation	Autosome
DPOGS215098	CLIP-like	27161	Modulation	Autosome
DPOGS208169	CLIP-like	3911	Modulation	Autosome
DPOGS201966	CLIP-like	10337	Modulation	Autosome
DPOGS215182	CLIP-like	10004	Modulation	Autosome
DPOGS211355	CLIP-like	3601	Modulation	Autosome
DPOGS203664	CLIP-like	7166	Modulation	Autosome
DPOGS210568	CLIP-like	8283	Modulation	Autosome
DPOGS214570	CLIP-like	4377	Modulation	Autosome
DPOGS204148	CLIP-like	2905	Modulation	Autosome
DPOGS205210	CLIP-like	3342	Modulation	Autosome
DPOGS211237	CLIP-like	2943	Modulation	Autosome
DPOGS206224	CLIP-like	5720	Modulation	Autosome
DPOGS206217	CLIP-like	3300	Modulation	Autosome
DPOGS205206	CLIP-like	3310	Modulation	Autosome
DPOGS209809	SPZ-like	1777	Signaling - Toll	Autosome
DPOGS209810	SPZ-like	5018	Signaling - Toll	Autosome
DPOGS203200	Toll_like-receptors	3821	Signaling - Toll	Autosome
DPOGS205279	Toll_like-receptors	4664	Signaling - Toll	Autosome
DPOGS202626	Toll_like-receptors	2709	Signaling - Toll	Autosome
DPOGS205281	Toll_like-receptors	3894	Signaling - Toll	Autosome
DPOGS205295	Toll_like-receptors	3887	Signaling - Toll	Autosome
DPOGS205123	Toll_like-receptors	5897	Signaling - Toll	Autosome
DPOGS211472	Toll_like-receptors	4533	Signaling - Toll	Autosome
DPOGS200002	Toll_like-receptors	5428	Signaling - Toll	Autosome
DPOGS205283	Toll_like-receptors	1949	Signaling - Toll	Autosome
DPOGS215274	Toll_like-receptors	5667	Signaling - Toll	Autosome
DPOGS203198	Toll_like-receptors	3410	Signaling - Toll	Autosome
DPOGS205293	Toll_like-receptors	869	Signaling - Toll	Autosome
DPOGS205296	Toll_like-receptors	2006	Signaling - Toll	Autosome
DPOGS207788	Tollip	3828	Signaling - Toll	Autosome
DPOGS205936	MyD88	2965	Signaling - Toll	Autosome
DPOGS208945	Tube	2121	Signaling - Toll	Autosome
DPOGS214647	Pellino	3601	Signaling - Toll	Autosome
DPOGS210260	Pelle	11577	Signaling - Toll	Autosome
DPOGS202662	TRAF2	6521	Signaling - Toll	Autosome
DPOGS209243	ECSIT	1486	Signaling - Toll	Autosome

Gene ID	Gene name	Gene length	Functional class	Chromosome
DPOGS209453	Cactus	2181	Signaling - Toll	Autosome
DPOGS215778	IMD	1319	Signaling - IMD	Autosome
DPOGS200403	TAK1	12371	Signaling - IMD	Autosome
DPOGS202907	IKKgamma	14063	Signaling - IMD	Autosome
DPOGS202564	IKKbeta	1586	Signaling - IMD	Autosome
DPOGS207960	FADD	916	Signaling - IMD	Autosome
DPOGS212093	Dredd	1508	Signaling - IMD	Autosome
DPOGS200977	Tab2	3087	Signaling - IMD	Autosome
DPOGS203759	IAP2	5539	Signaling - IMD	Autosome
DPOGS201405	Ubc13	455	Signaling - IMD	Autosome
DPOGS208954	Hem	3377	Signaling - JNK	Autosome
DPOGS213169	JNK	3769	Signaling - JNK	Autosome
DPOGS214573	Fos	2706	Signaling - JNK	Autosome
DPOGS202887	Jun	242	Signaling - JNK	Autosome
DPOGS214325	PIAS	12079	Signaling - JAK-STAT	Autosome
DPOGS214451	SOCS	2617	Signaling - JAK-STAT	Autosome
DPOGS200349	DOMELESS	2960	Signaling - JAK-STAT	Autosome
DPOGS210157	Hopscotch	8831	Signaling - JAK-STAT	Autosome
DPOGS212956	Stat	16032	Signaling - JAK-STAT	Autosome
DPOGS213997	Attacin-Like	877	Effector	Autosome
DPOGS205720	Attacin-Like	1439	Effector	Autosome
DPOGS215451	Attacin-Like	818	Effector	Autosome
DPOGS210270	Cecropin-like	374	Effector	Autosome
DPOGS210268	Cecropin-like	422	Effector	Autosome
DPOGS210269	Cecropin-like	529	Effector	Autosome
DPOGS200256	Cecropin-like	428	Effector	Autosome
DPOGS210271	Cecropin-like	352	Effector	Autosome
DPOGS210265	Cecropin-like	848	Effector	Autosome
DPOGS210304	Gloverin-like	644	Effector	Autosome
DPOGS210303	Gloverin-like	822	Effector	Autosome
DPOGS202093	NOS-like	15626	Effector	Autosome
DPOGS202094	NOS-like	19723	Effector	Autosome
DPOGS201818	PPO-like	4660	Effector	Z-chromosome
DPOGS201819	PPO-like	5674	Effector	Z-chromosome
DPOGS206820	PPO-like	9352	Effector	Z-chromosome
DPOGS200017	PPO-like	5087	Effector	Autosome

Table S4.3. Population genetic statistics of immune genes in the North American population using the paired-control approach, based on all sites within each gene. The F_{ST} section was non-applicable because the North American population was the reference population used for population comparisons. "All immune" indicates the full immune gene set. In each statistic, the first raw shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % were labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 were labeled in bold. Asterisks indicates: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
All sites					
π: test statistic	-0.05	0.03	-0.17	0.05	0.03
π: > 0 (%)	22.49	90.47	**0.00**	94.55	84.97
π: *P*-value.	0.451	0.196	**0.000*** **	0.109	0.295
Watterson's θ: test statistic	-0.03	0.05	-0.19	0.08	0.03
Watterson's θ: > 0 (%)	34.28	94.35	**0.00**	**98.69**	79.51
Watterson's θ: *P*-value	0.687	0.122	**0.000*** **	**0.028***	0.414
Tajima's D: test statistic	-8.46	0.66	-11.63	-2.60	5.11
Tajima's D: > 0 (%)	4.09	63.89	**0.01**	12.66	**98.27**
Tajima's D: *P*-value	0.084	0.718	**0.001****	0.250	**0.030***
F_{ST}: test statistic	NA	NA	NA	NA	NA
F_{ST}: > 0 (%)	NA	NA	NA	NA	NA
F_{ST}: *P*-value	NA	NA	NA	NA	NA

Table S4.4. Population genetic statistics of immune genes in the Florida population using the paired-control approach, based on all sites within each gene. F_{ST} was compared to the North American population. "All immune" indicates the full immune gene set. In each statistic, the first raw shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % were labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 were labeled in bold. Asterisks indicates: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
All sites					
π: test statistic	-0.07	0.04	-0.17	0.05	0.01
π: > 0 (%)	13.75	92.24	**0.00**	94.41	59.58
π: *P*-value.	0.278	0.160	**0.000*** **	0.111	0.844
Watterson's θ: test statistic	-0.05	0.05	-0.18	0.07	0.01
Watterson's θ: > 0 (%)	23.95	94.95	**0.00**	**97.76**	57.51
Watterson's θ: *P*-value	0.479	0.108	**0.000*** **	**0.046***	0.866
Tajima's D: test statistic	-12.35	0.89	-13.48	-2.80	3.04
Tajima's D: > 0 (%)	**0.99**	67.38	**0.02**	9.40	87.62
Tajima's D: *P*-value	**0.021***	0.650	**0.000*** **	0.189	0.248
F_{ST}: test statistic	-0.09	0.04	-0.02	-0.06	-0.05
F_{ST}: > 0 (%)	35.52	75.72	48.40	34.08	25.14
F_{ST}: *P*-value	0.679	0.485	0.920	0.579	0.489

Table S4.5. Population genetic statistics of immune genes in the Pacific population using the paired-control approach, based on all sites within each gene. F_{ST} was compared to the North American population. "All immune" indicates the full immune gene set. In each statistic, the first raw shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % were labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 were labeled in bold. Asterisks indicates: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
All sites					
π: test statistic	-0.03	0.04	-0.14	0.04	0.04
π: > 0 (%)	29.57	97.12	**0.00**	92.75	86.69
π: *P*-value.	0.570	0.058	**0.000*** **	0.139	0.243
Watterson's θ: test statistic	-0.02	0.04	-0.12	0.04	0.03
Watterson's θ: > 0 (%)	32.68	**99.41**	**0.00**	95.93	87.26
Watterson's θ: *P*-value	0.635	**0.013***	**0.000*** **	0.077	0.231
Tajima's D: test statistic	-1.29	-2.94	2.26	-1.78	1.17
Tajima's D: > 0 (%)	43.76	16.54	68.12	32.22	62.68
Tajima's D: *P*-value	0.873	0.329	0.630	0.662	0.754
F_{ST}: test statistic	-0.50	-0.21	0.08	-0.43	0.06
F_{ST}: > 0 (%)	26.09	24.37	58.39	16.57	58.58
F_{ST}: *P*-value	0.505	0.461	0.862	0.327	0.878

Table S4.6. Population genetic statistics of immune genes in the Atlantic population using the paired-control approach, based on all sites within each gene. F_{ST} was compared to the North American population. "All immune" indicates the full immune gene set. In each statistic, the first raw shows the test statistic of the immune gene group. The second row shows the proportion of 10,000 permutations in which the difference between the means of the immune gene group and the control set was positive. Percentages < 2.5% and > 97.5 % were labeled in bold. The third row shows the *P*-value. *P*-values < 0.05 were labeled in bold. Asterisks indicates: * < 0.05, ** < 0.01, *** < 0.001.

	All Immune	**Recognition**	**Signaling**	**Modulation**	**Effector**
All sites					
π: test statistic	-0.04	0.00	-0.12	0.03	0.05
π: > 0 (%)	24.54	58.07	**0.00**	83.25	92.68
π: *P*-value.	0.483	0.852	**0.001****	0.324	0.103
Watterson's θ: test statistic	0.00	0.01	-0.10	0.04	0.04
Watterson's θ: > 0 (%)	47.51	78.50	**0.01**	93.64	93.63
Watterson's θ: *P*-value	0.940	0.447	**0.001****	0.122	0.087
Tajima's D: test statistic	-21.03	-6.60	-8.49	-6.55	0.61
Tajima's D: > 0 (%)	**1.53**	6.75	8.62	8.51	55.26
Tajima's D: *P*-value	**0.029***	0.119	0.167	0.159	0.880
F_{ST}: test statistic	1.16	0.30	1.37	0.07	-0.57
F_{ST}: > 0 (%)	91.75	81.55	**99.38**	57.80	3.48
F_{ST}: *P*-value	0.156	0.368	**0.009****	0.875	0.094

Table S4.7. Immune genes are not disproportionally represented in genome-wide outliers. We used chi-square tests to evaluate whether immune genes (n = 102) are disproportionally represented in genome-wide outliers in each population at either 0-fold or 4-fold degeneracy sites. Outliers were defined as < 2.5th percentile or > 97.5th percentile of the genome background in either Tajima's D or F_{ST}. For the North American population, outliers were identified based on only the Tajima's D data.

Population	Sites	Number of outlier immune genes	$\chi 2$	*df*	P-value
North America	0-fold	3	0.47	1	0.49
	4-fold	5	0.00	1	1.00
South Florida	0-fold	13	0.86	1	0.35
	4-fold	9	0.00	1	1.00
Pacific	0-fold	8	0.16	1	0.69
	4-fold	4	2.67	1	0.10
Atlantic	0-fold	9	0.01	1	0.91
	4-fold	13	0.59	1	0.44

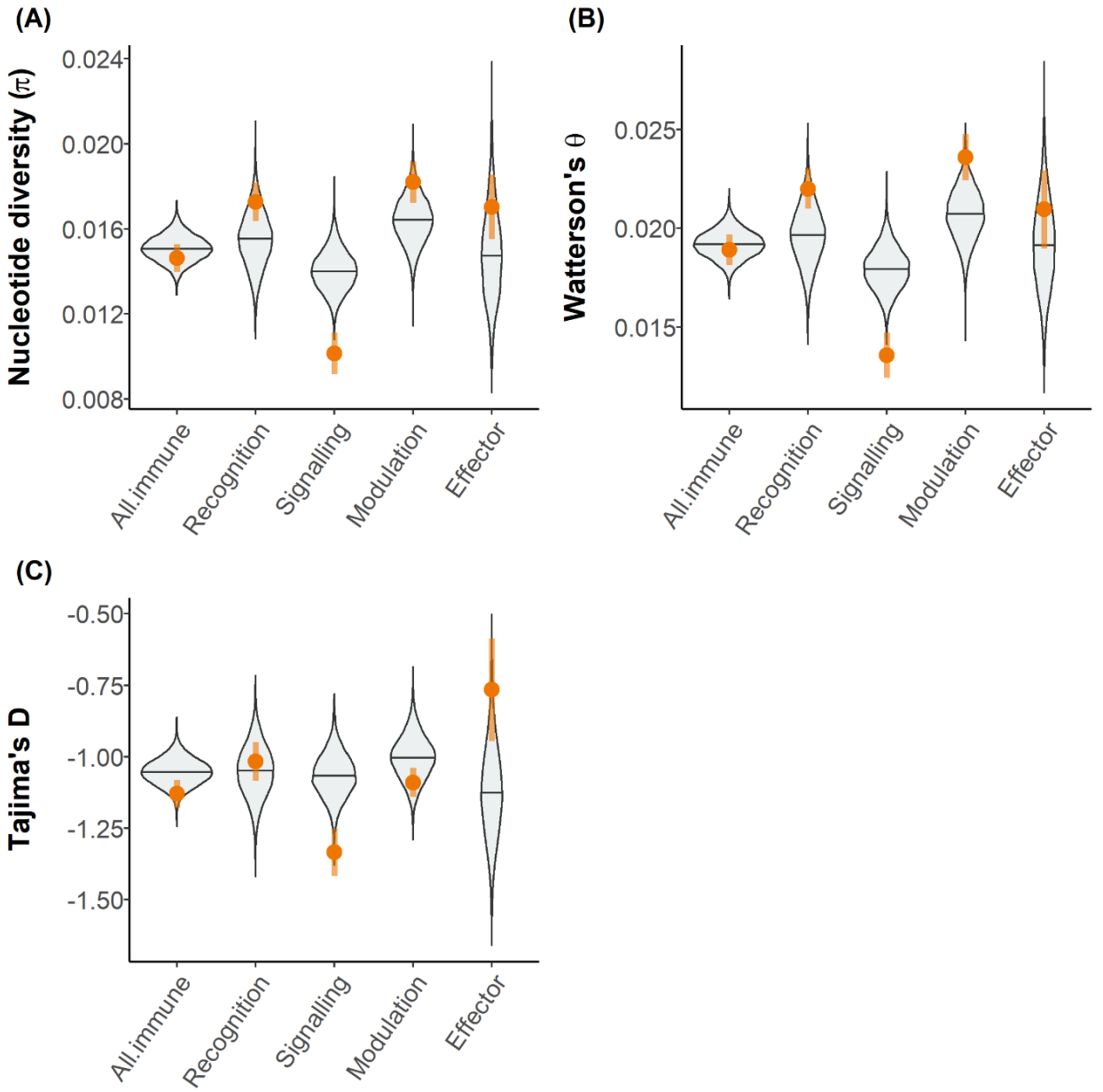


Figure S4.1. Population genetic statistics of immune genes in the North American population using the paired-control approach, based on all sites within each gene. (a): Nucleotide diversity (***π***); (b): Watterson's θ; (c): Tajima's D. Each immune gene group was compared to selected pair-control sets. Violin plots show the distribution of the mean of each control set generated with 10, 000 permutations. The orange dots and vertical lines indicate mean ±1 SEM of the immune gene group of interest.

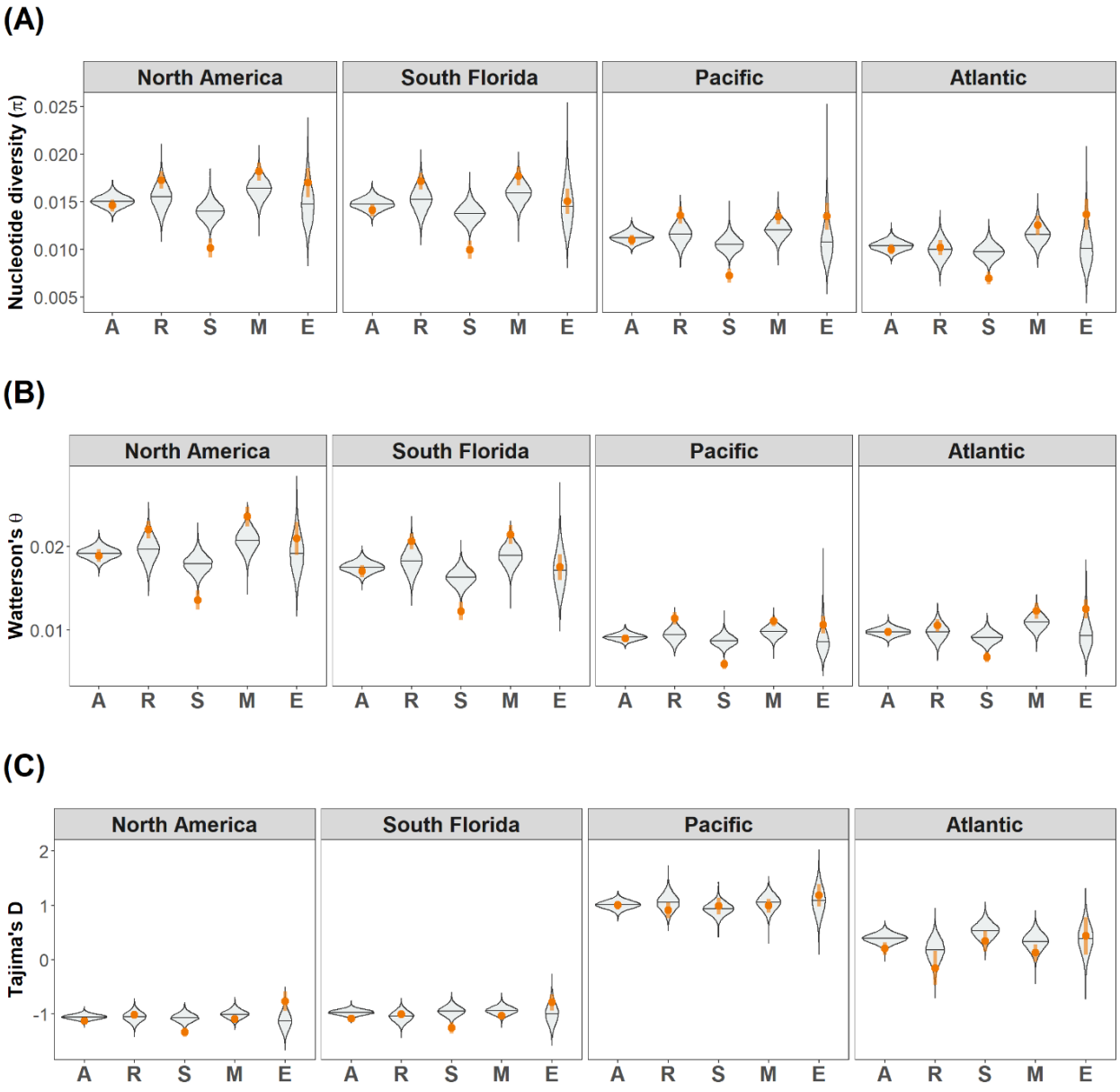


Figure S4.2. Population genetic statistics of immune genes in all four populations (North America, Florida, Pacific, and Atlantic) using the paired-control approach, based on all sites within each gene. (a): Nucleotide diversity (***π***); (b): Watterson's θ; (c): Tajima's D. Each immune gene group was compared to selected pair-control sets. Violin plots show the distribution of the mean of each control set generated with 10, 000 permutations. The orange dots and vertical lines indicate mean ±1 SEM of the immune gene group of interest. X-axis represents immune gene groups: all immune genes (A), recognition genes (R), signaling genes (S), modulation genes (M), and effector genes (E).

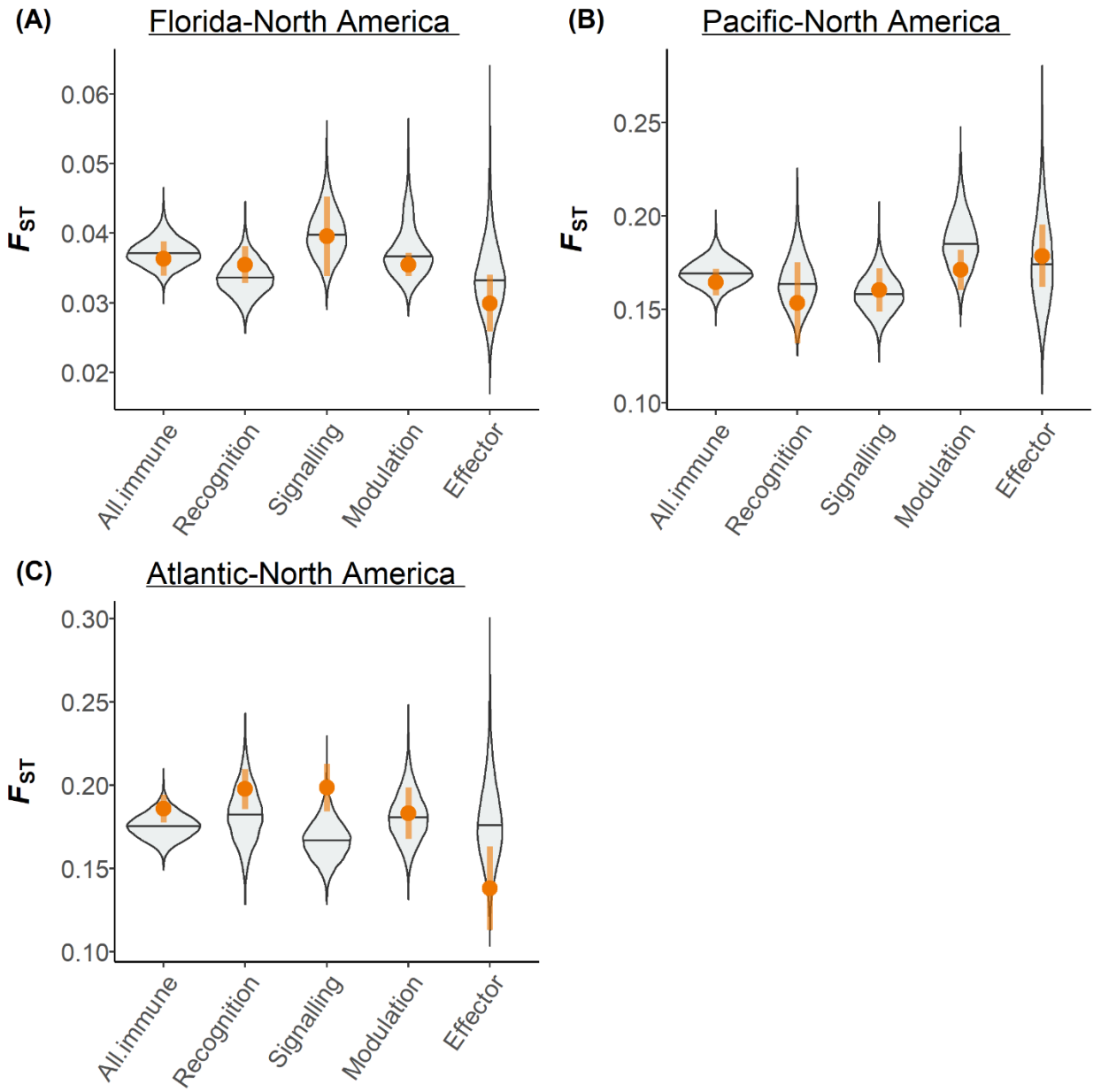


Figure S4.3. F_{ST} of immune genes in each derived population compared to the ancestral (North American) population using the paired-control approach, based on all sites within each gene. (a): South Florida population ($\boldsymbol{\pi}$); (b): Pacific population; (c): Atlantic population. Each immune gene group was compared to selected pair-control sets. Violin plots show the distribution of the mean of each control set generated with 10, 000 permutations. The orange dots and vertical lines indicate mean ±1 SEM of the immune gene group of interest.

Chapter 5: Conclusions and future directions

Defense is not only determined by the pairwise interaction between the host and parasite; the ecological context which the host and parasite interact also shapes host defense (Lazzaro & Little, 2009; Wolinska & King, 2009). The environmental context proximately can shape the ecological interactions across trophic levels, and ultimately can also shape the evolution of animal defenses. In the milkweed-monarch-parasite system, monarchs not only have multiple resistance mechanisms against milkweed defense chemicals (cardenolides) but also can sequester cardenolides for their own defense against natural enemies (Després et al., 2007). Monarchs feeding on high-cardenolide milkweed species have increased resistance and tolerance against parasites (Sternberg et al., 2012). In this dissertation, I studied monarch butterfly defenses in the tri-trophic interaction between milkweeds, monarchs, and parasites. In chapter 2, I studied how milkweed induction, which increased cardenolide production, affect the interactions between monarchs and parasites. Next, in chapter 3, I studied global gene expression of monarch larvae in response to milkweeds, parasites, and the potential interactions between the two. Lastly, in chapter 4, I studied the evolution of canonical immune genes across different monarch populations that vary in with their association with milkweed species and parasite prevalence. Here, I summarize the main findings and discuss future research directions for each chapter.

5.1 Discussion of Chapter 2

In this monarch-milkweed-parasite tri-trophic system, previous research has repeatedly shown that foliar cardenolides increase monarch resistance and tolerance to the parasite (*e.g.*, de Roode, Pedersen, Hunter, & Altizer, 2008; Sternberg, de Roode, & Hunter, 2015; Sternberg et al.,

2012; Tao, Gowler, Ahmad, Hunter, & de Roode, 2015). In this chapter, I tested for the effect of milkweed induction, which increases cardenolide production, on monarch resistance against the parasite. Our results showed that herbivory by monarch larvae induces a significant increase in foliar cardenolide concentrations in *A. curassavica* but not in *A. syriaca*. However, induction of cardenolides in *A. curassavica*, with a 1.3-fold increase in total concentration, did not translate into a statistically significant increase in monarch resistance against parasite infection. Our results suggest that interspecific variation in cardenolide concentration is a more important driver of parasite defense than plasticity via induction in this tri-trophic system.

Even though cardenolides were not induced in *A. syriaca* and parasite resistance was not affected, we observed reductions in the lifespan of monarchs when fed with *A. syriaca*. This result suggests that induction altered other milkweed qualities aside from cardenolides, and that these changes affected monarch fitness. Therefore, future studies should measure an array of defense mechanisms and nutritional properties in response to herbivore damage to comprehensively evaluate the effects of induction on multi-trophic interactions.

Although the magnitude of cardenolide induction was not strong enough to significantly affect monarch resistance against the parasite, it is possible that the changes in cardenolide concentration due to induction might affect monarch oviposition choice. Previous studies have shown that females infected with parasites preferentially lay more eggs on high-cardenolide milkweeds (Lefèvre et al., 2012, 2010). Thus, it is possible that infected females may prefer herbivore-induced plants. Future study should examine the effect of cardenolide induction on female oviposition choice and the consequences of these choices for multi-trophic interactions. If infected females can distinguish between induced and non-induced plants, milkweed induced defense could indirectly alter density-dependent effects on monarchs and parasite transmission

dynamics.

5.2 Discussion of Chapter 3

Monarchs have multiple mechanisms to overcome milkweed chemical defenses, and feeding on highly toxic milkweeds also reduces parasite infection. Toxic plants could directly interfere with parasites or enhance immunity, or they could down-regulate immunity by providing an alternative defense against parasitism. In Chapter 3, I used RNA-Seq to assess differences in transcriptional profiles between monarch larvae feeding on the low-cardenolide *A. incarnata* and the high-cardenolide *A. curassavica* when infected or uninfected with the parasite. Our results demonstrate that monarch larvae differentially express several hundred genes when feeding on *A. curassavica* and *A. incarnata*. Given that the two milkweed species are similar in nutrient composition (Tao, Ahmad, et al., 2016) but differ greatly in their cardenolide concentration, these transcriptional differences are likely related to coping with different levels of toxicity in their diet. In fact, these differentially expressed genes include many that are from multiple families of canonical insect detoxification genes.

Regulation of gene expression in response to variable levels of toxicity may not only link to detoxification but also sequestration. Monarch larvae can regulate the level of cardenolide sequestration (Jones et al., 2019; Malcolm, 1991). Interestingly, our results show that all the differentially expressed ABC transporters were up-regulated in larvae fed *A. incarnata*, a milkweed species with very low cardenolide concentrations. ABC transporters play a key role in toxin sequestration in other insect systems (Strauss et al., 2013). Therefore, the up-regulation of ABC transporters when feeding on low-cardenolide milkweed might be related to an increased rate of cardenolide sequestration. However, relatively little is known about the mechanism of

cardenolide sequestration in insects (Jones et al., 2019; Petschenka & Agrawal, 2016), and follow-up studies are needed to verify the possible role of ABC transporters in cardenolide sequestration by monarchs.

We observed almost no transcriptional response to parasite infection regardless of host plant diet. It is possible that the parasite is more active and/or has a stronger interaction with the host immune system at different stages of the infection cycle. Thus, additional life stages should be taken into consideration in future studies. Nevertheless, paralogs of some of the differentially expressed genes between plant diet have functional relevance in anti-parasitic defenses in the mosquito-*Plasmodium* system. Future studies examining the functional roles of these particular immune genes using genome editing techniques such as the CRISPR/Cas system (Li et al., 2015; Markert et al., 2016) would be needed to verify their potential role in defense against *O. elektroscirrha* infections.

Insect hosts may reduce investment in costly immune responses when anti-parasite resistance is provided by host plants instead. Although we did not find a strong overall effect of plant diet on the expression of canonical immune genes, our results show that all four significantly differentially expressed canonical immune genes were down-regulated in monarchs fed with *A. curassavica*, the anti-parasitic plant species. This finding is consistent with the hypothesis that medicinal plants could reduce immune investment by providing an alternative form of anti-parasite defense (de Roode et al., 2013). Nevertheless, this does not preclude the possibility that other monarch immune defenses not captured by gene expression differences, may be influenced by host plant diet. Future studies should strive to characterize the function of the many genes of unknown function in monarchs, some of which could play a role in defense.

5.3 Discussion of Chapter 4

Parasite and pathogens exert strong selection pressures on host defense, so immune genes are presumably rapidly evolving (McTaggart et al., 2012), but the evolution of immune genes is also constrained by trade-offs with other biological functions and shaped by environmental context (Demas & Nelson, 2012). Monarchs originated in North America and colonized worldwide locations in the 19th century (Ackery & Vane-Wright, 1984). Through these dispersal events, monarchs formed populations in which they associate with different milkweed communities that vary in toxicity and in which they face different pathogen pressure, including facing differences in *O. elektroscirrha* prevalence (Altizer & de Roode, 2015). Chapter 2 confirmed previous studies showing that milkweed species is an important driver of parasite resistance, and chapter 3 showed that milkweed chemicals might have some interactions with immune gene expression. In this chapter, using a population genomic dataset, we examined the evolutionary patterns of the full set of known canonical immune genes within and among populations of monarch butterflies, and relative to a closely related species (*D. gilippus*). We found that immune genes as a whole were not under an uniform selection regime. Signaling genes exhibited consistently low levels of genetic variation across populations and between the two *Danaus* species, indicating they are likely highly constrained. While consistencies across populations are also found for modulation genes, they exhibited the opposite pattern - signatures of relaxed selection. In contrast, effector and recognition genes exhibited less consistent patterns across populations. In some populations, effector genes exhibited signatures of balancing selection, while recognition genes exhibited directional selection and population differentiation. When focusing on individual genes that are genomic outliers, we found some clear differences among populations. Variable patterns across populations are consistent with the hypothesis that the worldwide dispersal of monarchs resulted in differential

selection on some immune genes.

Notably, our results are consistent with the limited body of recent literature that examined population genetics of the full set of insect canonical immune genes (Early et al., 2017; Keehnen et al., 2018). The emergence of a common pattern across insect species that differ considerably in life histories and taxonomy also indicates that there may be some general evolutionary patterns among insect immune genes. Future studies on a wider range of insect taxa would be needed to confirm macroevolutionary patterns.

The identification of immune genes that are under differential selection in monarch populations opens the way for further functional characterization. Specifically, population-specific patterns indicate a possibility of local adaptation, and functional characterization is needed to understand the phenotypic effects of different alleles of immune genes, especially as they relate to important ecological factors. Future studies can use genome editing techniques such as the CRISPR/Cas system (Li et al., 2015; Markert et al., 2016) to evaluate the effect of these identified immune genes on monarch defense against an array of natural enemies. Beyond immune genes, future studies on monarch genomics should strive to improve the genomic resources of monarchs. For instance, a chromosome level re-assembly of the monarch butterfly reference genome, improved characterization of gene functions, and a fine-scale resequencing of individuals from each geographic population with more biological replicates, could enhance resolution and provide additional insight into the evolutionary patterns of monarchs.

Last but not least, although monarch butterflies are a charismatic organism and have been a research model in many biological contexts, such as migration, aposematism, and chemical ecology of plant-insect interactions, further natural history studies are still needed, especially for those non-American derived populations. Specifically, in some geographic populations, several

important ecological factors, including the community composition of milkweeds and their toxicity, hostplant preference and medication behavior, the prevalence of *O. elektroscirrha*, as well as other pathogen pressures, remain poorly understood. Altogether, future studies that characterize the natural histories of worldwide monarch populations, in combination with population genomic studies, would advance our knowledge of the evolutionary ecology of monarch defenses in a tri-trophic context.

Bibliography

Abbott, J. (2014). Self-medication in insects: Current evidence and future perspectives. *Ecological Entomology*, *39*, 273–280. doi:10.1111/een.12110

Ackery, P. R., & Vane-Wright, R. I. (1984). *Milkweed butterflies: their cladistics and biology*. Ithaca, NY: Cornell University Presss.

Agrawal, A. A. (2005). Natural selection on common milkweed (*Asclepias syriaca*) by a community of specialized insect herbivores. *Evolutionary Ecology Research*, *7*(5), 651–667. doi:10.1021/ct300848z

Agrawal, A. A., Hastings, A. P., Patrick, E. T., & Knight, A. C. (2014). Specificity of herbivore-induced hormonal signaling and defensive traits in five closely related milkweeds (*Asclepias* spp.). *Journal of Chemical Ecology*, *40*(7), 717–729. doi:10.1007/s10886-014-0449-6

Agrawal, A. A., Kearney, E. E., Hastings, A. P., & Ramsey, T. E. (2012). Attenuation of the jasmonate burst, plant defensive traits, and resistance to specialist monarch caterpillars on shaded common milkweed (*Asclepias syriaca*). *Journal of Chemical Ecology*, *38*(7), 893–901. doi:10.1007/s10886-012-0145-3

Agrawal, A. A., Petschenka, G., Bingham, R. A., Weber, M. G., & Rasmann, S. (2012). Toxic cardenolides: chemical ecology and coevolution of specialized plant-herbivore interactions. *New Phytologist*, *194*(1), 28–45. doi:10.1111/j.1469-8137.2011.04049.x

Altizer, S., & de Roode, J. C. (2015). Monarchs and their debilitating parasites: immunity, migration, and

medicinal plant use. In *Monarchs in a Changing World*.

Altizer, S., Hobson, K. A., Davis, A. K., De Roode, J. C., & Wassenaar, L. I. (2015). Do healthy monarchs migrate farther? Tracking natal origins of parasitized vs. Uninfected monarch butterflies overwintering in Mexico. *PLoS ONE*, *10*(11), e0141371. doi:10.1371/journal.pone.0141371

Altizer, S., Oberhauser, K. S., & Brower, L. P. (2000). Associations between host migration and the prevalence of a protozoan parasite in natural populations of adult monarch butterflies. *Ecological Entomology*, *25*, 125–139. doi:10.1046/j.1365-2311.2000.00246.x

Altizer, S., Oberhauser, K. S., & Geurts, K. A. (2004). Transmission of the protozoan parasite *Ophryocystis elektroscirrha* in monarch butterfly populations: implications for prevalence and population-level impacts. In *The monarch butterfly: biology and conservation* (pp. 203–218).

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

Barillas-Mury, C. (2007). CLIP proteases and *Plasmodium* melanization in *Anopheles gambiae*. *Trends in Parasitology*, *23*(7), 297–299. doi:10.1016/j.pt.2007.05.001

Bartel, R. A., Oberhauser, K. S., de Roode, J. C., & Altizer, S. (2011). Monarch butterfly migration and parasite transmission in eastern North America. *Ecology*, *92*(2), 342–351. doi:10.1890/10-0489.1

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society . Series B (Methodological)*,

57(1), 289–300.

Birnbaum, S. S., Rinker, D. C., Gerardo, N. M., & Abbot, P. (2017). Transcriptional profile and differential fitness in a specialist milkweed insect across host plants varying in toxicity. *Molecular Ecology*, *26*(23), 6742–6761. doi:10.1111/mec.14401

Blumberg, B. J., Trop, S., Das, S., & Dimopoulos, G. (2013). Bacteria- and IMD pathway-independent immune defenses against *Plasmodium falciparum* in *Anopheles gambiae*. *PLoS ONE*, *8*(9), e72130. doi:10.1371/journal.pone.0072130

Bowers, M. D. (1992). The evolution of unpalatability and the cost of chemical defense in insects. In M. B. Roitberg, B.D. & Isman (Ed.), *Insect Chemical Ecology: An Evolutionary Approach* (pp. 216–244). New York: Chapman & Hall.

Brower, L. P., McEvoy, P. B., Williamson, K. L., & Flannery, M. A. (1972). Variation in cardiac glycoside content of monarch butterflies from natural populations in eastern North America. *Science*, *177*(4047), 5–8. doi:10.1126/science.177.4047.426

Brower, L. P., Ryerson, W. N., Coppinger, L. L., & Susan, C. (1968). Ecological chemistry and the palatability spectrum. *Science*, *161*(3848), 1349–1351.

Carlson, M., & Pages, H. (2018). AnnotationForge: code for building annotation database packages.

Castella, G., Chapuisat, M., Moret, Y., & Christe, P. (2008). The presence of conifer resin decreases the use of the immune system in wood ants. *Ecological Entomology*, *33*(3), 408–412. doi:10.1111/j.1365-

2311.2007.00983.x

Castellano, D., Coronado-Zamora, M., Campos, J. L., Barbadilla, A., & Eyre-Walker, A. (2016). Adaptive evolution is substantially impeded by hill-Robertson interference in *Drosophila*. *Molecular Biology and Evolution*, *33*(2), 442–455. doi:10.1093/molbev/msv236

Chapman, J. R., Hill, T., & Unckless, R. L. (2018). Balancing selection drives maintenance of genetic variation in *Drosophila* antimicrobial peptides. *bioRxiv*, 298893. doi:10.1101/298893

Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., … Kafatos, F. C. (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science*, *298*(5591), 159–165. doi:10.1126/science.1077136

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. doi:10.4161/fly.19695

Comeron, J. M., Ratnappan, R., & Bailin, S. (2012). The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics*, *8*(10), e1002905. doi:10.1371/journal.pgen.1002905

Connallon, T., & Clark, A. G. (2015). The distribution of fitness effects in an uncertain world. *Evolution*, *69*(6), 1610–1618. doi:10.1111/evo.12673

Cory, J. S., & Hoover, K. (2006). Plant-mediated effects in insect-pathogen interactions. *Trends in Ecology and Evolution*, *21*(5), 278–286. doi:10.1016/j.tree.2006.02.005

Crava, C. M., Brütting, C., & Baldwin, I. T. (2016). Transcriptome profiling reveals differential gene expression of detoxification enzymes in a hemimetabolous tobacco pest after feeding on jasmonate-silenced Nicotiana attenuata plants. *BMC Genomics*, *17*(1), 1005. doi:10.1186/s12864-016-3348-0

Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., … Jiggins, C. D. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94–98. doi:10.1038/nature11041

de la Paz Celorio-Mancera, M., Wheat, C. W., Vogel, H., Söderlind, L., Janz, N., & Nylin, S. (2013). Mechanisms of macroevolution: Polyphagous plasticity in butterfly larvae revealed by RNA-Seq. *Molecular Ecology*, *22*(19), 4884–4895. doi:10.1111/mec.12440

de Roode, J. C., Chi, J., Rarick, R. M., & Altizer, S. (2009). Strength in numbers: High parasite burdens increase transmission of a protozoan parasite of monarch butterflies (*Danaus plexippus*). *Oecologia*, *161*(1), 67–75. doi:10.1007/s00442-009-1361-6

de Roode, J. C., Fernandez de, C. L., Faits, T., & Alizon, S. (2011). Virulence evolution in response to anti-infection resistance: Toxic food plants can select for virulent parasites of monarch butterflies. *Journal of Evolutionary Biology*, *24*(4), 712–722. doi:10.1111/j.1420-9101.2010.02213.x

de Roode, J. C., Lefèvre, T., & Hunter, M. D. (2013). Self-medication in animals. *Science*, *340*(6129), 150–151. doi:10.1126/science.1235824

de Roode, J. C., Pedersen, A. B., Hunter, M. D., & Altizer, S. (2008). Host plant species affects virulence

in monarch butterfly parasites. *Journal of Animal Ecology*, *77*(1), 120–126. doi:10.1111/j.1365-2656.2007.01305.x

de Roode, J. C., Rarick, R. M., Mongue, A. J., Gerardo, N. M., & Hunter, M. D. (2011). Aphids indirectly increase virulence and transmission potential of a monarch butterfly parasite by reducing defensive chemistry of a shared food plant. *Ecology Letters*, *14*(5), 453–461. doi:10.1111/j.1461-0248.2011.01604.x

de Roode, J. C., Yates, A. J., Altizer, S., & Roode, J. C. De. (2008). Virulence-transmission trade-offs and population divergence in virulence in a naturally occurring butterfly parasite. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(21), 7489–7494. doi:10.1073/pnas.0710909105

Demas, G. E., & Nelson, R. J. (2012). *Ecoimmunology*. New York: Oxford University Press.

Després, L., David, J. P., & Gallet, C. (2007). The evolutionary ecology of insect resistance to plant chemicals. *Trends in Ecology & Evolution*, *22*(6), 298–307. doi:10.1016/j.tree.2007.02.010

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. doi:10.1093/bioinformatics/bts635

Dobler, S., Dalla, S., Wagschal, V., & Agrawal, A. A. (2012). Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na , K-ATPase. *Proceedings of the*

National Academy of Sciences, *109*(32), 13040–13045. doi:10.1073/pnas.1202111109

Dong, Y., & Dimopoulos, G. (2009). *Anopheles* fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *Journal of Biological Chemistry*, *284*(15), 9835–9844. doi:10.1074/jbc.M807084200

Dong, Y., Simões, M. L., Marois, E., & Dimopoulos, G. (2018). CRISPR/Cas9 -mediated gene knockout of Anopheles gambiae FREP1 suppresses malaria parasite infection. *PLoS Pathogens*, *14*(3), e1006898. doi:10.1371/journal.ppat.1006898

Dyer, L. A., & Bowers, M. D. (1996). The importance of sequestered iridoid glycosides as a defense against an ant predator. *Journal of Chemical Ecology*, *22*(8), 1527–1539. doi:10.1007/BF02027729

Early, A. M., Arguello, J. R., Cardoso-Moreira, M., Gottipati, S., Grenier, J. K., & Clark, A. G. (2017). Survey of global genetic diversity within the *Drosophila* immune system. *Genetics*, *205*(1), 353–366. doi:10.1534/genetics.116.195016

Early, A. M., & Clark, A. G. (2017). Genomic signatures of local adaptation in the *Drosophila* immune response. *Fly*, *11*(4), 277–283. doi:10.1080/19336934.2017.1337612

Eizaguirre, C., Lenz, T. L., Kalbe, M., & Milinski, M. (2012). Divergent selection on locally adapted major histocompatibility complex immune genes experimentally proven in the field. *Ecology Letters*, *15*(7), 723–731. doi:10.1111/j.1461-0248.2012.01791.x

Erler, S., Lhomme, P., Rasmont, P., & Lattorff, H. M. G. (2014). Rapid evolution of antimicrobial peptide

genes in an insect host-social parasite system. *Infection, Genetics and Evolution*, *23*, 129–137. doi:10.1016/j.meegid.2014.02.002

Estévez-Lao, T. Y., & Hillyer, J. F. (2014). Involvement of the *Anopheles gambiae* Nimrod gene family in mosquito immune responses. *Insect Biochemistry and Molecular Biology*, *44*(1), 12–22. doi:10.1016/j.ibmb.2013.10.008

Evans, J. D., Aronstein, K., Chen, Y. P., Hetru, C., Imler, J. L., Jiang, H., … Hultmark, D. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Molecular Biology*, *15*(5), 645–656.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. doi:10.1093/bioinformatics/btw354

Eyre-Walker, A., & Keightley, P. D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution*, *26*(9), 2097–2108. doi:10.1093/molbev/msp119

Fordyce, J. A. (2001). The lethal plant defense paradox remains: Inducible host-plant aristolochic acids and the growth and defense of the pipevine swallowtail. *Entomologia Experimentalis et Applicata*, *100*(3), 339–346. doi:10.1023/A:1019249306992

Foster, S. P., Bass, C., Williamson, M. S., Millar, N. S., Field, L. M., Oliphant, L., … Puinean, A. M. (2010).

Amplification of a cytochrome P450 gene is associated with resistance to neonicotinoid insecticides in the aphid *Myzus persicae*. *PLoS Genetics*, *6*(6), e1000999. doi:10.1371/journal.pgen.1000999

Gerardo, N. M., Altincicek, B., Anselme, C., Atamian, H., Barribeau, S. M., de Vos, M., … Vilcinskas, A. (2010). Immunity and other defenses in pea aphids, *Acyrthosiphon pisum*. *Genome Biology*, *11*(2), R21. doi:10.1186/gb-2010-11-2-r21

Glick, B., Day, E. J., & Thompson, D. (1981). Calorie-protein deficiencies and the immune response of the chicken I. humoral immunity. *Poultry Science*, *60*(11), 2494–2500.

Glick, B., Robert L. Taylor, J., Martin, D. E., Watabe, M., Day, E. J., & Thompson, D. (1983). Calorie-protein deficiencies and the immune response of the chicken.: II. Cell-mediated immunity. *Poultry Science*, *62*(9), 1889–1893.

Gowler, C. D., Leon, K. E., Hunter, M. D., & de Roode, J. C. (2015). Secondary defense chemicals in milkweed reduce parasite infection in monarch butterflies, *Danaus plexippus*. *Journal of Chemical Ecology*, *41*(6), 520–523. doi:10.1007/s10886-015-0586-6

Gurung, P., & Kanneganti, T. D. (2015). Innate immunity against *Leishmania* infections. *Cellular Microbiology*, *17*(9), 1286–1294. doi:10.1111/cmi.12484

Hoang, K. L., Matzkin, L. M., & Bono, J. M. (2015). Transcriptional variation associated with cactus host plant adaptation in *Drosophila mettleri* populations. *Molecular Ecology*, *24*(20), 5186–5199. doi:10.1111/mec.13388

Hunter, M. D., & Schultz, J. C. (1993). Induced plant defenses breached ? Phytochemical induction protects an herbivore from disease. *Oecologia*, *94*(2), 195–203.

Jones, P. L., Peschenka, G., Flacht, L., & Agrawal, A. A. (2019). Cardenolide intake, sequestration, and excretion by the monarch butterfly along gradients of plant toxicity and larval ontogeny. *Journal of Chemical Ecology*. doi:10.1007/s10886-019-01055-7

Karban, R., & Baldwin, I. T. (1997). *Induced responses to herbivory*. Chicago: University of Chicago Press.

Keehnen, N. L. P., Hill, J., Nylin, S., & Wheat, C. W. (2018). Microevolutionary selection dynamics acting on immune genes of the green-veined white butterfly, *Pieris napi*. *Molecular Ecology*, *27*(13), 2807–2822. doi:10.1111/mec.14722

Kim, Y. S., Ryu, J. H., Han, S. J., Choi, K. H., Nam, K. B., Jang, I. H., … Lee, W. J. (2000). Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and β-1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *Journal of Biological Chemistry*, *275*(42), 32721–32727. doi:10.1074/jbc.M003934200

Korneliussen, T. S., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356. doi:doi:10.1186/s12859-014-0356-4

Kurucz, É., Márkus, R., Zsámboki, J., Folkl-Medzihradszky, K., Darula, Z., Vilmos, P., … Andó, I. (2007). Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Current Biology*, *17*(7), 649–654. doi:10.1016/j.cub.2007.02.041

Lampert, E. C. (2012). Influences of plant traits on immune responses of specialist and generalist herbivores. *Insects*, *3*(2), 573–592. doi:10.3390/insects3020573

Lampert, E. C., & Bowers, M. D. (2015). Incompatibility between plant-derived defensive chemistry and immune response of two Sphingid herbivores. *Journal of Chemical Ecology*, *41*(1), 85–92. doi:10.1007/s10886-014-0532-z

Lande, R. (2009). Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, *22*(7), 1435–1446. doi:10.1111/j.1420-9101.2009.01754.x

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–9. doi:10.1038/nmeth.1923

Laurentz, M., Reudler, J. H., Mappes, J., Friman, V., Ikonen, S., & Lindstedt, C. (2012). Diet quality can play a critical role in defense efficacy against parasitoids and pathogens in the glanville fritillary (*Melitaea cinxia*). *Journal of Chemical Ecology*, *38*(1), 116–125. doi:10.1007/s10886-012-0066-1

Lazzaro, B. P., & Little, T. J. (2009). Immunity in a variable world. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1513), 15–26. doi:10.1098/rstb.2008.0141

Lee, K. P., Simpson, S. J., & Wilson, K. (2008). Dietary protein-quality influences melanization and immune function in an insect. *Functional Ecology*, *22*(6), 1052–1061. doi:10.1111/j.1365-2435.2008.01459.x

Lefèvre, T., Chiang, A., Kelavkar, M., Li, H., Li, J., de Castillejo, C. L. F., … de Roode, J. C. (2012). Behavioural resistance against a protozoan parasite in the monarch butterfly. *Journal of Animal Ecology*, *81*(1), 70–79. doi:10.1111/j.1365-2656.2011.01901.x

Lefèvre, T., Oliver, L., Hunter, M. D., & de Roode, J. C. (2010). Evidence for trans-generational medication in nature. *Ecology Letters*, *13*(12), 1485–1493. doi:10.1111/j.1461-0248.2010.01537.x

Leighton, T., Marks, E., & Leighton, F. (1981). Pesticides : insecticides and fungicides are chitin synthesis inhibitors. *Science*, *213*(4510), 905–907.

Lemaitre, B., & Hoffmann, J. A. (2007). The host defense of *Drosophila melanogaster*. *Annual Review of Immunology*, *25*(1), 697–743. doi:10.1146/annurev.immunol.25.022106.141615

Leong, K. L. H., Yoshimura, M. A., & Kaya, H. K. (1997). Occurrence of a neogregarine protozoan, *Ophryocystis elektroscirrha* McLaughlin and Myers, in populations of monarch and queen butterflies. *The Pan-Pacific Entomologist*, *73*, 49–51.

Li, X., Fan, D., Zhang, W., Liu, G., Zhang, L., Zhao, L., … Wang, W. (2015). Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. *Nature Communications*, *6*, 1–10. doi:10.1038/ncomms9212

Little, T. J., Hultmark, D., & Read, A. F. (2005). Invertebrate immunity and the limits of mechanistic immunology. *Nature Immunology*, *6*(7), 651–654. doi:10.1038/ni1219

Lunter, G., & Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of

Illumina sequence reads. *Genome Research*, *21*(3), 936–939. doi:10.1101/gr.111120.110 [doi]

MacGregor, P., Szöőr, B., Savill, N. J., & Matthews, K. R. (2012). Trypanosomal immune evasion, chronicity and transmission: An elegant balancing act. *Nature Reviews Microbiology*, *10*(6), 431–438. doi:10.1038/nrmicro2779

Malcolm, S. B. (1991). Cardenolide-mediated interactions between platns and herbivores. In G. Rosenthal & M. Berenbaum (Eds.), *Herbivores: Their interactions with secondary plant metabolites, Vol. I: The chemical participants* (second edi, pp. 251–291). San Diego.

Malcolm, S. B. (1994). Milkweeds, monarch butterflies and the ecological significance of cardenolides. *Chemoecology*, *5*(3–4), 101–117. doi:10.1007/BF01240595

Mao, W., Rupasinghe, S., Zangerl, A. R., Schuler, M. A., & Berenbaum, M. R. (2006). Remarkable substrate-specificity of CYP6AB3 in *Depressaria pastinacella*, a highly specialized caterpillar. *Insect Molecular Biology*, *15*(2), 169–179. doi:10.1111/j.1365-2583.2006.00623.x

Markert, M. J., Zhang, Y., Enuameh, M. S., Reppert, S. M., Wolfe, S. A., & Merlin, C. (2016). Genomic access to monarch migration using TALEN and CRISPR/Cas9-mediated targeted mutagenesis. *G3: Genes, Genomes, Genetics*, *6*(4), 905–915. doi:10.1534/g3.116.027029

Marty, M. A., & Krieger, R. I. (1984). Metabolism of uscharidin, a milkweed cardenolide, by tissue homogenates of monarch butterfly larvae, *Danaus plexippus L. Journal of Chemical Ecology*, *10*(6), 945–956.

Matskevich, A. A., Quintin, J., & Ferrandon, D. (2010). The *Drosophila* PRR GNBP3 assembles effector complexes involved in antifungal defenses independently of its Toll-pathway activation function. *European Journal of Immunology*, *40*(5), 1244–1254. doi:10.1002/eji.200940164

Matzkin, L. M. (2012). Population transcriptomics of cactus host shifts in *Drosophila mojavensis*. *Molecular Ecology*, *21*(10), 2428–2439. doi:10.1111/j.1365-294X.2012.05549.x

McKenna, A., Hanna, M., Banks, E., Sivachenko, Andrey Cibulskis, K., Kernytsky, A., Garimella, K., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. doi:doi:10.1101/gr.107524.110

Mclaughlin, R. E., & Myers, J. (1970). *Ophrypcystis elektroscirrha* sp.., a Neogregarine pathogen of monarch butterfly *Danaus plexippus* (L .) and the Florida queen butterfly *D* . *gilippus* berenice Cramer. *Journal of Protozoology*, *17*(2), 300–305. doi:10.1111/j.1550-7408.1970.tb02375.x

McTaggart, S. J., Obbard, D. J., Conlon, C., & Little, T. J. (2012). Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. *BMC Evolutionary Biology*, *12*(1), 63. doi:10.1186/1471-2148-12-63

Miner, B. G., Sultan, S. E., Morgan, S. G., Padilla, D. K., & Relyea, R. A. (2005). Ecological consequences of phenotypic plasticity. *Trends in Ecology and Evolution*. doi:10.1016/j.tree.2005.08.002

Mongue, A. J., Hansen, M. E., Gu, L., Sorenson, C. E., & Walters, J. R. (2019). Non-fertilizing sperm in Lepidoptera show little evidence for recurrent positive selection. *Molecular Ecology, In press*.

doi:10.1111/mec.15096

Mongue, A. J., Nguyen, P., Volenikova, A., & Walters, J. R. (2017). Neo-sex chromosomes in the monarch butterfly, *Danaus plexippus*. *G3: Genes, Genomes, Genetics*, *7*(October), g3.300187.2017. doi:10.1534/g3.117.300187

Mooney, K. A., Jones, P. L., & Agrawal, A. A. (2008). Coexisting congeners: Demography, competition, and interactions with cardenolides for two milkweed-feeding aphids. *Oikos*, *117*(3), 450–458. doi:10.1111/j.2007.0030-1299.16284.x

Moret, Y., & Schmid-Hempel, P. (2000). Survival for immunity: the price of immune system activation for bumblebee workers. *Science*, *290*(November), 1166–1169. doi:10.1126/science.290.5494.1166

Muller, K., Vogelweith, F., Thiéry, D., Moret, Y., & Moreau, J. (2015). Immune benefits from alternative host plants could maintain polyphagy in a phytophagous insect. *Oecologia*, *177*(2), 467–475. doi:10.1007/s00442-014-3097-1

Nei, M. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10), 5269–5273. doi:10.1073/pnas.76.10.5269

Nielsen, R., & Slatkin, M. (2013). *An introduction to population genetics: theory and applications*. Massachusetts: Sinauer Associates, Inc.

Nishida, R. (2002). Sequestration of defense substances from plants by Lepidoptera. *Annual Review of Entomology*, *47*, 57–92. doi:10.1146/annurev.ento.47.091201.145121

Oberhauser, K. S., & Solensky, M. J. (2004). *The monarch butterfly: biology and conservation*. Ithaca, NY: Cornell University Presss.

Ojala, K., Julkunen-Tiitto, R., Lindström, L., & Mappes, J. (2005). Diet affects the immune defence and life-history traits of an Arctiid moth *Parasemia plantaginis*. *Evolutionary Ecology Research*, *7*(8), 1153–1170.

Oliver, K. M., Russell, J. A., Moran, N. A., & Hunter, M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, *100*(4), 1803–1807. doi:10.1073/pnas.0335320100

Opitz, S. E. W., & Müller, C. (2009). Plant chemistry and insect sequestration. *Chemoecology*, *19*(3), 117–154. doi:10.1007/s00049-009-0018-6

Parker, B. J., Barribeau, S. M., Laughton, A. M., de Roode, J. C., & Gerardo, N. M. (2011). Non-immunological defense in an evolutionary framework. *Trends in Ecology and Evolution*, *26*(5), 242–248. doi:10.1016/j.tree.2011.02.005

Parker, B. J., Spragg, C. J., Altincicek, B., & Gerardo, N. M. (2013). Symbiont-mediated protection against fungal pathogens in pea aphids: A role for pathogen specificity. *Applied and Environmental Microbiology*, *79*(7), 2455–2458. doi:10.1128/AEM.03193-12

Petschenka, G., & Agrawal, A. A. (2015). Milkweed butterfly resistance to plant toxins is linked to sequestration, not coping with a toxic diet. *Proceedings of the Royal Society B: Biological Sciences*,

282(1818), 20151865. doi:10.1098/rspb.2015.1865

Petschenka, G., & Agrawal, A. A. (2016). How herbivores coopt plant defenses: natural selection, specialization, and sequestration. *Current Opinion in Insect Science*, *14*, 17–24. doi:10.1016/j.cois.2015.12.004

Petschenka, G., Offe, J. K., & Dobler, S. (2012). Physiological screening for target site insensitivity and localization of Na +/K +-ATPase in cardenolide-adapted Lepidoptera. *Journal of Insect Physiology*, *58*(5), 607–612. doi:10.1016/j.jinsphys.2011.12.012

Pierce, A. A., Zalucki, M. P., Bangura, M., Udawatta, M., Kronforst, M. R., Altizer, S., … de Roode, J. C. (2014). Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20142230. doi:10.1098/rspb.2014.2230

Quintero, C., Lampert, E. C., & Bowers, M. D. (2014). Time is of the essence: direct and indirect effects of plant ontogenetic trajectories on higher trophic levels. *Ecology*, *95*(9), 2589–2602.

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

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

Rasmann, S., Johnson, M. D., & Agrawal, A. A. (2009). Induced responses to herbivory and jasmonate in

three milkweed species. *Journal of Chemical Ecology*, *35*(11), 1326–1334. doi:10.1007/s10886-009-9719-0

Richards, L. A., Lampert, E. C., Bowers, M. D., Dodson, C. D., Smilanich, A. M., & Dyer, L. A. (2012). Synergistic effects of iridoid glycosides on the survival, development and immune response of a specialist caterpillar, *Junonia coenia* (Nymphalidae). *Journal of Chemical Ecology*, *38*(10), 1276–1284. doi:10.1007/s10886-012-0190-y

Ricklefs, R. E., & Wikelski, M. (2002). The physiology/life-history nexus. *Trends in Ecology & Evolution*, *17*(10), 462–468.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). EdgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. doi:10.1093/bioinformatics/btp616

Rosenthal, G., & Berenbaum, M. R. (1991). *Herbivores: their interaction with secondary plant metabolites, Vol. II: Ecological and evolutionary processes* (second edi). San Diego: Academic Press.

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., & Clark, A. G. (2007). Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genetics*, *39*(12), 1461–1468. doi:10.1038/ng.2007.60

Sadd, B. M., & Schmid-hempel, P. (2009). A distinct infection cost associated with trans-generational priming of antibacterial immunity in bumble-bees. *Biology Letters*, *5*(6), 798–801.

doi:10.1098/rsbl.2009.0458

Schlenke, T. A., & Begun, D. J. (2003). Natural selection drives *Drosophila* immune system evolution. *Genetics*, *164*(4), 1471–1480.

Schoonhoven, L., van Loon, J., & Dicke, M. (2005). *Insect-plant biology* (second edi). New York: Oxford University Press.

Schuler, M. A. (1996). The role of cytochrome P450 monooxygenases in plant-insect interactions. *Plant Physiology*, (112), 1411–1419.

Selkirk, M. E., Bundy, D. A. P., Smith, D. F., Anderson, R. M., & Maizels, R. M. (2003). Immunological modulation and evasion by helminth parasites in human populations. *Nature*, *365*(6449), 797–805. doi:10.1038/365797a0

Sheldon, B. C., & Verhulst, S. (1996). Ecological immunology: Costly parasite defences and trade-offs in evolutionary ecology. *Trends in Ecology and Evolution*, *11*(8), 317–321. doi:10.1016/0169-5347(96)10039-2

Simões, M. L., Dong, Y., Hammond, A., Hall, A., Crisanti, A., Nolan, T., & Dimopoulos, G. (2017). The *Anopheles* FBN9 immune factor mediates *Plasmodium* species-specific defense through transgenic fat body expression. *Developmental and Comparative Immunology*, *67*, 257–265. doi:10.1016/j.dci.2016.09.012

Simone, M., Evans, J. D., & Spivak, M. (2009). Resin collection and social immunity in honey bees.

Evolution, *63*(11), 3016–3022. doi:10.1111/j.1558-5646.2009.00772.x

Singer, M. S., Mace, K. C., & Bernays, E. A. (2009). Self-medication as adaptive plasticity: Increased ingestion of plant toxins by parasitized caterpillars. *PLoS ONE*, *4*(3), e4796. doi:10.1371/journal.pone.0004796

Smilanich, A. M., Dyer, L. A., Chambers, J. Q., & Bowers, M. D. (2009). Immunological cost of chemical defence and the evolution of herbivore diet breadth. *Ecology Letters*, *12*(7), 612–621. doi:10.1111/j.1461-0248.2009.01309.x

Smilanich, A. M., Dyer, L. A., & Gentry, G. L. (2014). The insect immune response and other putative defenses as effective predictors of parasitism. *Ecology*, *90*(6), 1434–1440.

Smilanich, A. M., Teglas, M. B., Harrison, J. G., Hsueh, J., Dyer, L. A., Langus, T. C., & Doan, L. (2017). Host plant associated enhancement of immunity and survival in virus infected caterpillars. *Journal of Invertebrate Pathology*, *151*(October 2017), 102–112. doi:10.1016/j.jip.2017.11.006

Smilanich, A. M., Vargas, J., Dyer, L. A., & Bowers, M. D. (2011). Effects of ingested secondary metabolites on the immune response of a polyphagous caterpillar *Grammia incorrupta*. *Journal of Chemical Ecology*, *37*(3), 239–245. doi:10.1007/s10886-011-9924-5

Somogyi, K., Sipos, B., Pénzes, Z., & Andó, I. (2010). A conserved gene cluster as a putative functional unit in insect innate immunity. *FEBS Letters*, *584*(21), 4375–4378. doi:10.1016/j.febslet.2010.10.014

Stearns, S. . C. . (1989). Trade-offs in life-history evolution. *Functional Ecology*, *3*(3), 259–268.

Sternberg, E. D., de Roode, J. C., & Hunter, M. D. (2015). Trans-generational parasite protection associated with paternal diet. *Journal of Animal Ecology*, *84*(1), 310–321. doi:10.1111/1365-2656.12289

Sternberg, E. D., Lefèvre, T., Li, J., Lopez, C., Castillejo, F. De, Li, H., … Roode, J. C. De. (2012). Food plant-derived disease tolerance and resistance in a natural butterfly-plant-parasite interactions. *Evolution*, *66*(11), 3367–3377. doi:10.5061/dryad.82j66

Strauss, A. S., Peters, S., Boland, W., & Burse, A. (2013). ABC transporter functions as a pacemaker for sequestration of plant glucosides in leaf beetles. *eLife*, *2013*(2), 1–16. doi:10.7554/eLife.01096

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585–595.

Tan, W.-H., Tao, L., Hoang, K. M., Hunter, M. D., & de Roode, J. C. (2018). The effects of milkweed induced defense on parasite resistance in monarch butterflies, *Danaus plexippus*. *Journal of Chemical Ecology*, *44*(11), 1040–1044. doi:10.1007/s10886-018-1007-4

Tanji, T., Hu, X., Weber, A. N. R., & Ip, Y. T. (2007). Toll and IMD pathways synergistically activate an innate immune response in *Drosophila melanogaster*. *Molecular and Cellular Biology*, *27*(12), 4578–4588. doi:10.1128/MCB.01814-06

Tao, L., Ahmad, A., de Roode, J. C., & Hunter, M. D. (2016). Arbuscular mycorrhizal fungi affect plant tolerance and chemical defences to herbivory through different mechanisms. *Journal of Ecology*, *104*(2), 561–571. doi:10.1111/1365-2745.12535

Tao, L., Gowler, C. D., Ahmad, A., Hunter, M. D., & de Roode, J. C. (2015). Disease ecology across soil boundaries: effects of below-ground fungi on above-ground host–parasite interactions. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1817), 20151993. doi:10.1098/rspb.2015.1993

Tao, L., Hoang, K. M., Hunter, M. D., & de Roode, J. C. (2016). Fitness costs of animal medication: antiparasitic plant chemicals reduce fitness of monarch butterfly hosts. *The Journal of Animal Ecology*, *85*(5), 1246–1254. doi:10.1111/1365-2656.12558

Tataru, P., Mollion, M., Glémin, S., & Bataillon, T. (2017). Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *Genetics*, *107*(3), 1103–1119. doi:https://doi.org/10.1534/genetics.117.300323

Tauszig-Delamasure, S., Bilak, H., Capovilla, M., Hoffmann, J. A., & Imler, J. L. (2002). *Drosophila* MyD88 is required for the response to fungal and Gram-positive bacterial infections. *Nature Immunology*, *3*(1), 91–97. doi:10.1038/ni747

Theodoratus, D. H., & Bowers, M. D. (1999). Effects of sequestered iridoid glycosides on prey choice of the prairie wolf spider, *Lycosa carolinensis*. *Journal of Chemical Ecology*, *25*(2), 283–295. doi:10.1023/A:1020894729188

Törönen, P., Medlar, A., & Holm, L. (2018). PANNZER2: A rapid functional annotation web server. *Nucleic Acids Research*, *46*(W1), W84–W88. doi:10.1093/nar/gky350

Torres-Dowdall, J., Handelsman, C. A., Reznick, D. N., & Ghalambor, C. K. (2012). Local adaptation and

the evolution of phenotypic plasticity in trinidadian guppies (*Poecilia Reticulata*). *Evolution*, *66*(11), 3432–3443. doi:10.1111/j.1558-5646.2012.01694.x

Uésugi, K. (2010). The adaptive significance of Batesian mimicry in the swallowtail butterfly, *Papilio polytes* (Insecta, Papilionidae): associative learning in a predator. *Ethology*, *102*(5), 762–775. doi:10.1111/j.1439-0310.1996.tb01165.x

Unckless, R. L., Howick, V. M., & Lazzaro, B. P. (2016). Convergent balancing selection on an antimicrobial peptide in *Drosophila*. *Current Biology*, *26*(2), 257–262. doi:10.1016/j.cub.2015.11.063

Unckless, R. L., & Lazzaro, B. P. (2016). The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1695), 20150291. doi:10.1098/rstb.2015.0291

Viljakainen, L., Evans, J. D., Hasselmann, M., Rueppell, O., Tingek, S., & Pamilo, P. (2009). Rapid evolution of immune proteins in social insects. *Molecular Biology and Evolution*, *26*(8), 1791–1801. doi:10.1093/molbev/msp086

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *Annu. Rev. Genet*, *47*, 97–120. doi:10.1146/annurev-genet-111212-133526

Vlachou, D., Schlegelmilch, T., Christophides, G. K., & Kafatos, F. C. (2005). Functional genomic analysis of midgut epithelial responses in *Anopheles* during *Plasmodium* invasion. *Current Biology*, *15*(13), 1185–1195. doi:10.1016/j.cub.2005.06.044

Vogel, H., Musser, R. O., & de la Paz Celorio-Mancera, M. (2014). Transcriptome responses in herbivorous insects towards host plant and toxin feeding. In *Annual Plant Reviews: Insect-Plant Interactions* (Vol. 47, pp. 197–233). doi:10.1002/9781118829783.ch6

Volz, J., Müller, H. M., Zdanowicz, A., Kafatos, F. C., & Osta, M. A. (2006). A genetic module regulates the melanization response of *Anopheles* to *Plasmodium*. *Cellular Microbiology*, *8*(9), 1392–1405. doi:10.1111/j.1462-5822.2006.00718.x

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R., Huber, W., Liaw, A., … Venables, B. (2016). gplots: Various R programming tools for plotting data. Retrieved from https://cran.r-project.org/package=gplots

Waterhouse, R. M., Kriventseva, E. V., Meister, S., Xi, Z., Alvarez, K. S., Bartholomay, L. C., … Christophides, G. K. (2007). Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science*, *316*(5832), 1738–1743. doi:10.1126/science.1139862

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*(2), 256–276. doi:10.1016/0040-5809(75)90020-9

Wen, Z., Pan, L., Berenbaum, M. R., & Schuler, M. A. (2003). Metabolism of linear and angular furanocoumarins by *Papilio polyxenes* CYP6B1 co-expressed with NADPH cytochrome P450 reductase. *Insect Biochemistry and Molecular Biology*, *33*, 937–947. doi:10.1016/S0965-1748(03)00100-0

Werck-Reichhart, D., & Feyereisen, R. (2000). Cytochromes P450: a success story. *Genome Biology*, *1*(6), reviews3003.1-3003.9. doi:10.1186/gb-2000-1-6-reviews3003

Wolinska, J., & King, K. C. (2009). Environment can alter selection in host-parasite interactions. *Trends in Parasitology*, *25*(5), 236–244. doi:10.1016/j.pt.2009.02.004

Wong, A., Singh, N. D., Sackton, T. B., Zhang, Y., Oliver, B., Greenberg, A. J., … Sturgill, D. (2008). Evolution of protein-coding genes in *Drosophila*. *Trends in Genetics*, *24*(3), 114–123. doi:10.1016/j.tig.2007.12.001

Wright, S. (1921). Systems of mating IV. The effects of selection. *Genetics*, *6*(2), 162–166.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. doi:10.1089/omi.2011.0118

Yu, Q.-Y., Fang, S.-M., Zhang, Z., & Jiggins, C. D. (2016). The transcriptome response of *Heliconius melpomene* larvae to a novel host plant. *Molecular Ecology*, *25*(19), 4850–4865. doi:10.1111/mec.13826

Zalucki, M. P., & Brower, L. P. (1992). Survival of first instar larvae of *Danaus plexippus* (Lepidoptera: Danainae) in relation to cardiac glycoside and latex content of *Asclepias humistrata* (Asclepiadaceae). *Chemoecology*, *3*(2), 81–93. doi:10.1007/BF01245886

Zalucki, M. P., Brower, L. P., & Alonso-M, A. (2001). Detrimental effects of latex and cardiac glycosides

on survival and growth of first-instar monarch butterfly larvae *Danaus plexippus* feeding on the sandhill milkweed *Asclepias humistrata*. *Ecological Entomology*, *26*(2), 212–224. doi:10.1046/j.1365-2311.2001.00313.x

Zalucki, M. P., Brower, L. P., & Malcolm, S. B. (1990). Oviposition by *Danaus plexippus* in relation to cardenolide content of three *Asclepias* species in the southeastern U.S.A. *Ecological Entomology*, *15*, 231–240.

Zhan, S., Merlin, C., Boore, J. L., & Reppert, S. M. (2011). The monarch butterfly genome yields insights into long-distance migration. *Cell*, *147*(5), 1171–1185. doi:10.1016/j.cell.2011.09.052

Zhan, S., & Reppert, S. M. (2013). MonarchBase: The monarch butterfly genome database. *Nucleic Acids Research*, *41*(D1), 758–763. doi:10.1093/nar/gks1057

Zhan, S., Zhang, W., Niitepõld, K., Hsu, J., Haeger, J. F., Zalucki, M. P., … Kronforst, M. R. (2014). The genetics of monarch butterfly migration and warning colouration. *Nature*, *514*(7522), 317–21. doi:10.1038/nature13812

Zhang, G., Niu, G., Franca, C. M., Dong, Y., Wang, X., Butler, N. S., … Li, J. (2015). *Anopheles* midgut FREP1 mediates *Plasmodium invasion*. *Journal of Biological Chemistry*, *290*(27), 16490–16501. doi:10.1074/jbc.M114.623165

Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., & Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community. *Science*, *337*(6102), 1634–1637. doi:10.1126/science.1226630.