CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

Genes involucrados en la especificidad de hospedero de la mariposa monarca

Tesis que presenta

L.C.G. Pablo Manuel González De la Rosa

Para Obtener el Grado de

Maestro en Ciencias

En la Especialidad de

Biología Integrativa

Directores de la Tesis:

Dr. Cei Abreu Goodger

Dra. Therese Ann Markow

Sinodales:

Dra. Selene L. Fernandez-Valverde

Dr. Rubén Rellán Álvarez

Irapuato, Guanajuato Noviembre, 2018

El presente trabajo fue realizado en el laboratorio de Genómica de Biodiversificación, bajo la dirección de la Dra. Therese Ann Markow, y Genómica Computacional del ARN, bajo la dirección del Dr. Cei Abreu Goodger, ambos laboratorios de la Unidad de Genómica Avanzada del CINVESTAV Irapuato. El trabajo se realizó con financiamiento brindado por la Alianza Carlos Slim.

AKNOWLEDGEMENTS

I thank my family for encouraging me to continue my studies and for being there always that I needed. I thank both my advisors, Cei and Teri, for guiding me. They prepared me beyond the project. I thank all the LBC for the good talks and help. Special thanks to Francisco Falcon for insightful discussions which helped me out when I could not see beyond my nose. Thanks to Atenea who inspired me to make every day an adventure. Although with less time to share, I thank Isaac and Carlos for involving me into a new circle of friends. Thanks to Teri's lab for receiving me with open arms. Special thanks to Mariana Ramírez who lay down this project and congenially initiated me in the project. And last but not least, I thank CONACyT for the scholarship (598430) which allowed me to pursue this Master's degree.

AGRADECIMIENTOS

Agradezco a mi familia por animarme a continuar mis estudios y por estar allí siempre que lo necesitaba. Agradezco a mis dos asesores, Cei y Teri, por guiarme. Me prepararon más allá del proyecto. Agradezco a todos los LBC por las buenas charlas y ayuda. Un agradecimiento especial a Francisco Falcón por las discusiones profundas que me ayudaron cuando no podía ver más allá de mi nariz. Gracias a Atenea que me inspiró para hacer de cada día una aventura. Aunque con menos tiempo para compartir, agradezco a Isaac y Carlos por involucrarme en un nuevo círculo de amigos. Gracias al laboratorio de Teri por recibirme con los brazos abiertos. Un agradecimiento especial a Mariana Ramírez, quien estableció este proyecto y me inició de manera amable en el proyecto. Y agradezco a CONACyT por la beca (598430) que me permitió obtener este título de maestría.

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

SUMMARY

Herbivory is a predominant strategy among insects and most of them have a narrowly delimited range of hosts. These specialized insects have developed behavioral and biochemical countermeasures to cope with the plant's defenses. The monarch butterfly, besides its charismatic appearance and unique migration, has been a model of study for chemical ecologists for more than 70 years due to the monarchs' ability to sequester toxic cardenolides from its hosts, the milkweeds. Several decades ago, the monarch's resistance to cardenolides was explained by at least one amino acid substitution. However, the sequestering mechanism is still poorly understood and even less is known about which other genes are relevant for the monarch-milkweed interaction. Recent studies have identified ABC transporters as critical components for sequestering allelochemicals. Furthermore, there are no previous studies of the role of microRNAs in plant-herbivore interactions. Here I study the protein-coding and microRNA genes that vary between monarch caterpillars after they feed for 24 hours on three different milkweeds: *Asclepias curassavica*, *Asclepias linaria* and *Gomphocarpus physocarpus*. The first two are natives to North America while the latter is a member of the African clade. As expected, due to the phylogenetic distances, larvae fed on *G. physocarpus* exhibited the greatest number of gene expression differences relative to larvae fed on either of the *Asclepia*s. However, these differentially expressed protein-coding and miRNA genes suggest that larvae fed on this host had a stunted growth possibly due to nutrient restriction. These changes were accompanied by the overexpression of an unexpectedly large proportion of ABC transporter genes. On the other hand, metabolism and detoxification genes were predominant among the genes that differed between larvae fed on the two *Asclepias*. Altogether, this work enriches the genomic resource of the monarch butterfly, postulates miR-278 as growth regulator in response to nutrient availability by targeting multiple genes of the Hippo signaling pathway and identified several ABC transporter genes which might be critical for the monarch's cardenolide sequestering mechanism.

Abbreviations

AC *Asclepias curassavica*

AL *Asclepias linaria*

GP *Gomphocarpus physocarpus*

FDR False Discovery Rate

RESUMEN

Los insectos, el grupo más diverso de animales, son en su mayoría herbívoros y de éstos la mayoría se especializa en un rango limitado de plantas. La especialización de los insectos herbívoros se ve reflejada en las contramedidas bioquímicas y de comportamiento que usan para evadir las defensas de las plantas. La mariposa monarca, además de su vistosa apariencia y excepcional migración, ha sido estudiada por ecólogos químicos por más de 70 años debido a su capacidad de secuestrar cardenólidos tóxicos a partir de las plantas en las que se especializó, los algodoncillos. Desde hace varias décadas se sabe que las monarcas son inmunes a la toxicidad de los cardenólidos debido a al menos una sustitución aminoacídica. No obstante, aún se desconoce el mecanismo con el que las monarcas secuestran estos químicos y aún menos se conoce sobre qué otros genes son relevantes para la interacción entre las monarcas y los algodoncillos. Recientemente se encontró que algunos transportadores dependientes de adenosina trifosfato (ABC) participan en secuestrar aleloquímicos. Por otro lado, aún no se ha estudiado si algún micro-ARN juega un papel en la interacción entre insectos herbívoros y sus correspondientes plantas anfitrionas. Aquí estudio qué genes cambian su expresión en orugas de mariposa monarca cuando son alimentadas con tres distintos algodoncillos: *Asclepias curassavica*, *Asclepias linaria* y *Gomphocarpus physocarpus*. De estas plantas anfitrionas, las dos primeras son nativas de Norteamérica mientras que la tercera pertenece a un clado africano. Como era de esperarse tan sólo por la distancia filogenética, las orugas alimentadas con *G. physocarpus* tuvieron perfiles de expresión mucho más diferentes respecto a las orugas alimentadas en algodoncillos del género *Asclepias*. Los cambios en el perfil de expresión, tanto de genes codificantes de proteínas como los de micro-ARNs, sugieren que las orugas tuvieron un desarrollo pobre al alimentarse con *G. physocarpus* debido a que incorporaron menos nutrientes. Los transportadores ABC sobresalen entre los genes diferencialmente expresados entre las *Asclepias* y *G. physocarpus*. Múltiples genes de detoxificación y metabolismo de carbohidratos se encuentran entre los pocos genes diferencialmente expresados entre las orugas alimentadas con las distintas *Asclepias*. En su conjunto, este trabajo enriquece los recursos genómicos de la mariposa monarca, postula el impacto miR-278 en el crecimiento de las orugas a través de la regulación de varios genes de la ruta de señalización Hippo e identifica varios transportadores ABC que podrían ser críticos en el mecanismo usado por la monarca para secuestrar cardenólidos.

INTRODUCTION

Herbivorous insects

One of the broadest quests in biology is to understand how species arise and diversify. Plants and phytophagous insects have been common objects of study for such enquiries given that they amount to a large fraction of eukaryotic diversity and these taxa themselves are very diverse (Schoonhoven, Van Loon, van Loon, & Dicke, 2005). Furthermore, herbivorous insects have a tight coevolutionary relationship with plants because they depend on plants for feeding, mating, ovipositing and habitat during their life cycle. Most herbivorous insects are specialized on one host family (Forister et al., 2015). Plants, on the other hand, may take advantage of insects (e.g. for pollination), but they also need to defend themselves against many phytophagous insects. Herbivore success depends on their ability to cope with the plant phenology, nutrient composition and physicochemical defenses. At the same time, community composition and diversity of plants can be driven by phytophagous insects (Endara et al., 2017). This co-evolutionary arms race leads both sides of this competition into further specialization (Ehrlich & Raven, 1964).

Herbivorous insects can be classified according to diet breadth. Insects feeding on several hosts from different families are considered generalists or polyphagous while other, more restricted feeders, are classified as oligophagous or even monophagous. Furthermore, host range is genetically determined (Wiklund, 1974). The clear delineation of host range provides a wide diversity of coevolutionary ecological interaction case studies. One hypothesis explaining evolution of phytophagous insects states that plants evolve chemical defenses that keep most predators at bay, but a narrow range of herbivores adapt to these special defenses and thus radiate to newly unoccupied niches (Ehrlich & Raven, 1964). Since the beginning of the last century, secondary metabolism has been thought to play a role in these interactions (Brues, 1920; Dethier, 1954; Ehrlich & Raven, 1964). Subsequently, plant secondary metabolites, those with no direct effect on plant growth, often have been implicated in defense against herbivores (Futuyma & Agrawal, 2009).

Plant defenses and insect countermeasures

Plant defenses can be classified as active if they are inducible or as static if they are constitutive (Gatehouse, 2002). Defenses are also classified as direct or indirect. If they attract natural enemies (e.g. parasitoids or predators) of the herbivores, they are indirect defenses (e.g. plant volatiles and nectar rewards) (A. Kessler & Baldwin, 2001; André Kessler & Baldwin, 2002). Direct defenses, on the other hand, include physical barriers such as trichomes, leaf shape and toughness (Howe & Jander, 2008). Chemical defenses, allelochemicals, can be classified as terpenoids, alkaloids and phenolics. Defensive proteins, such as proteinase inhibitors (Green & Ryan, 1972), are among direct biochemical inducible defenses. Although the combination of direct and indirect defenses confers protection against many phytophagous arthropods, some can counteract these defenses (Schoonhoven et al., 2005).

Insect counter-defenses can broadly be grouped either as behavioral or biochemical (Brattsten, 1988). Examples of behavioral counter-defenses are leaf trenching (i.e. the insect avoids cutting the leaf veins) and vein-cutting (Dussourd, 1999). They can also evade or reduce plant defenses through salivary secretions, by regurgitating or by taking advantage of microbial partners (Felton et al., 2014). However, most molecular studies of host-insect interaction have focused on post-ingestion processes (Vogel, Musser, & Celorio-Mancera, 2014). Insects counteract plant biochemical defenses through two complementary strategies: target site insensitive and detoxification mechanisms. The first is obtained through amino acid substitutions that confer reduced sensitivity to allelochemicals (Dobler, Dalla, Wagschal, & Agrawal, 2012; Ujvari et al., 2015; Zhen, Aardema, Medina, Schumer, & Andolfatto, 2012). The second strategy is further divided in three phases: I) decreasing their toxicity by reduction, hydrolysis or oxidation of substrate, II) facilitating their mobilization by conjugating it to other molecules; and III) elimination or transportation (Brattsten, 1988; Després, David, & Gallet, 2007). Furthermore, insects can acquire detoxification mechanisms through their gut microbes (Henry et al., 2013; Sugio, Dubreuil, Giron, & Simon, 2015).

Cytochrome P450 monooxygenases (P450s) and carboxylesterases (COEs) participate in phase I. Glutathione S-transferases (GSTs) and UDP-glucuronosyl transferases (UGTs) take part in the second phase. Finally, in the third phase,

the resulting water-soluble compounds can be transported through membranes with the aid of p-glycoprotein efflux carriers and ATP-binding cassette transporters (Bretschneider, Heckel, & Vogel, 2016). Because particular allelochemicals are already water soluble and some are sequestered instead of excreted, detoxification of some xenobiotics does not require all three phases.

ABC transporters can transport a wide variety of ligands ranging from small organic or inorganic molecules up to peptides, lipids, oligonucleotides and polysaccharides (Wilkens, 2015). ABC transporters participate in a wide array of processes besides detoxification and are amply conserved between species (Dassa, 2011). Regardless of function, they have been divided into eight subfamilies (from A to H) according to sequence similarity (Dermauw & Van Leeuwen, 2014). In particular, subfamilies ABCB and ABCC include proteins associated with drug resistance, also known as multidrug resistance (MDR) proteins (Allikmets, Gerrard, Hutchinson, & Dean, 1996; Sodani, Patel, Kathawala, & Chen, 2012).

Additionally, some insects not only neutralize or avoid the toxic effect of allelochemicals but even sequester them presumably as protection against predators. Only few studies have probed which genes are involved in sequestration across the insect gut (Petschenka & Agrawal, 2016). The first gene identified as possibly involved in insect sequestration was an ABC transporter in leaf beetles (*Chrysomela populi*) which aids in funneling secondary metabolites from the hemolymph into defensive glands (Strauss et al., 2014; Strauss, Peters, Boland, & Burse, 2013). More recently, knock-outs in *Drosophila melanogaster* have shown that multidrug and organic anion transporters can counter-defend cardenolide noxious effects (Groen et al., 2017).

Transcriptomic responses of insects to different hosts

Although specialized counter-defenses of some insects have been thoroughly characterized at the molecular level, an increasing number of host-herbivore interactions have been studied by assessing their genome-wide expression changes after feeding on different hosts (Simon et al., 2015; Vogel et al., 2014). Several studies reveal that detoxification and altered lipid and carbohydrate metabolism are frequent responses of insects to host shifts (Simon et al., 2015).

In the generalist cotton bollworm (*Helicoverpa armigera*), diets supplemented with higher gossypol (an allelochemical produced by some plants from the Malvacea family) concentrations resulted in increased expression of P450 monooxygenases CYP6AE14 and CYP6AE11 as well as UGTs and GSTs mainly in the midgut (Celorio-Mancera, Ahn, Vogel, & Heckel, 2011). Genes involved in primary, secondary and xenobiotic metabolism, together with environmental information processing and extracellular matrix-receptor pathways, were differentially expressed in *H. armigera* fed on different hosts and even on different plant structures of the same host (de la Paz Celorio-Mancera, Heckel, & Vogel, 2012). Genes of the polyphagous Swedish comma (*Polygonia c-album*) encoding serine-type endopeptidases, membrane-associated, transporters and nucleic acid binding proteins, were differentially expressed between larvae fed on common nettle (*Urtica dioica*) and gooseberry (*Ribes uva-crispa*) (De La Paz Celorio-Mancera et al., 2013). Koenig *et al.*, (2015) assessed the transcriptomic response of the specialist *M. sexta* larvae to feeding on different Solenaceous hosts and a Brassicaceous non-host plant, rapeseed (*Brassica napus*). Cytochrome P450 monooxygenases, UGTs, GSTs and ABC transporter detoxification genes were differentially expressed when larvae were fed for eight days on the different hosts. Late instar larvae of the postman butterfly (*Heliconius melpomene*) had different expression of genes encoding GSTs UGTs, components of peritrophic matrix and transporters, if they were fed on a native host, *Passiflora menispermifolia*, as opposed to a less well defended non-host, *Passiflora biflora* (Q. Y. Yu, Fang, Zhang, & Jiggins, 2016). Additionally, Bretschneider, Heckel and Vogel (2016) explored the host specific transcriptomic response of the cotton bollworm by supplementing the larval artificial diet with the secondary metabolites characteristic of different host plants (nicotine, taxol and tomatine). They identified a negative correlation between larval growth rate and number of differentially expressed genes which suggest that larvae invest in detoxification mechanisms at the expense of development. They propose several ABC transporters as candidates for detoxifying nicotine, tomatine and taxol. Combining genome wide associations and transcriptome profiling, Nallu *et al.*, (2018) recently studied which plant and insect genes might be relevant for host-herbivore interaction among *Papilio polytes*, *Colias eurytheme*, *Pieris rapae* and *Heliconius cydno* with their respective hosts *Citrofortunella microcarpa*, *Medicago*

sativa, *Arabidopsis thaliana* and *Passiflora oerstedii*. They found that only few orthologous genes are common responses, in the four lepidopterans, to their respective hosts. They argue that this lack of generality is reasonable if each herbivore-host interaction followed different trajectories of co-evolution. This suggests that genes relevant for one host-lepidopteran interaction might not be extrapolatable even to relatively phylogenetically close organisms. However, they only measured transcriptomic differences after interacting for a period of time but did not exposed the larvae to different hosts.

Roles of microRNAs in insects

Although much research has focused on genes encoding proteins, non-coding RNAs are also essential for a wide variety of biological processes (Cech & Steitz, 2014). In particular, microRNAs (miRNAs), RNA molecules of approximately 22 nucleotides discovered in *C. elegans* (R. C. Lee, Feinbaum, & Ambros, 1993; Wightman, Ha, & Ruvkun, 1993), can regulate almost 40% of *D. melanogaster* protein coding genes affecting nearly all cellular pathways (Agarwal, Subtelny, Thiru, Ulitsky, & Bartel, 2018). These molecules recruit the RNA-induced silencing complex (RISC) to posttranscriptionally repress their targets by partial complementarity against the mRNA's three prime untranslated region (3' UTR) (Bartel, 2018). Due to usually perfect Watson-Crick pairing between the miRNA 'seed' region (nucleotides 2-8) and complementary sites in 3' UTRs, it is possible to computationally predict miRNA targets at a genome wide level (Friedman, Farh, Burge, & Bartel, 2009; Lewis, Shih, Jones-Rhoades, Bartel, & Burge, 2003). However, these predictions include many false positives due to the random occurrence of complementary sites given how short the pairing is.

Conserved and non-conserved miRNAs can play critical roles in many insect's essential processes such as growth, morphogenesis, metabolism, behavior, and host-pathogen interactions among others (Lucas, Zhao, Liu, & Raikhel, 2015). Bantam, the first miRNA characterized in an insect, caused decreased adult *D. melanogaster* size when knocked-out but overgrowth in wings and eyes, when overexpressed, by negatively regulating the proapoptotic gene *hid* (Brennecke, Hipfner, Stark, Russell, & Cohen, 2003; Hipfner, Weigmann, & Cohen, 2002). Also in *D. melanogaster*, miR-2 proapoptotic targets, *reaper* and *grim*, were validated through observing diminished green fluorescent protein, fused with the

corresponding 3' UTR, where miR-2b was detected (Stark, Brennecke, Russell, & Cohen, 2003). Later, Leaman *et al.*, (2005) found that inhibiting miR-2/13 by injecting antisense 2′O-methyl oligoribonucleotides in *D. melanogaster* embryos increased the number of apoptotic cells already by stage 13, but this phenotype was rescued by constitutive overexpression of miR-2. However, when the injection of this antisense oligoribonucleotide was unable to trigger apoptosis in embryos deficient for the *hid*, *grim* or *reaper* genes (Leaman et al., 2005). On the other hand inhibiting miR-2 family with locked nucleic acids rendered the German cockroach (*Blatella germanica)* unable to transit from nymph to adult because this miRNA was unable to downregulate Krüppel-homolog 1 (Lozano, Montañez, & Belles, 2015). In contrast to the predictable relevance of widely conserved miRNAs, inhibition of miR-2942, only reported in the Asian tiger mosquito (*Aedes albopictus*), caused a decreased rate of egg hatching and larva eclosion (Puthiyakunnon et al., 2013). Hence, conserved and non-conserved miRNAs could be involved in the host-herbivore coevolutionary relationship.

The monarch butterfly and the milkweeds

The iconic monarch butterfly (*Danaus plexippus*) of North America is best known for its massive annual international migration that takes more than four generations to complete. Additionally, this charismatic butterfly has been a model in chemical ecology for more than 70 years thanks to the unpalatability against vertebrate predators that it acquires while feeding on a poisonous plant. However, the population of migratory monarch butterflies has drastically decreased over the last two decades due to habitat loss (Flockhart et al., 2013; Thogmartin et al., 2017; Vidal, López-García, & Rendón-Salinas, 2014). Monarch caterpillars only feed on plants colloquially known as milkweeds, which are easily recognizable by their milky exudation when their leaves are cut. It has been estimated that monarch reproduction in the Midwestern United States decreased by 81% between 1999 and 2010 mainly due to loss of milkweeds (Pleasants & Oberhauser, 2013; Semmens et al., 2016). Due to this environmental change and the phenological and geographical differences, monarchs might be forced to oviposit and develop in suboptimal hosts (Dingle, Zalucki, Rochester, & Armijo-Prewitt, 2005). Therefore, it is increasingly relevant to study how monarchs interact with different potential hosts.

The sticky latex that characterizes milkweeds can by itself be lethal to early instar caterpillars (M. P. Zalucki, Brower, & Alonso-M, 2001; Myron P Zalucki & Malcolm, 1999). Milkweeds are members of the Apocynacea family, primarily from the genus *Asclepias* (L. P. Brower, 1969). Plants of this family include among their defenses the production of cardiac glycosides which are particularly abundant in latex (Agrawal & Konno, 2009; Agrawal, Petschenka, Bingham, Weber, & Rasmann, 2012). Cardenolides disrupt secondary active transport and nerve cell membrane potential (Jorgensen, Håkansson, & Karlish, 2003) by binding to the α-subunit of the Na^+/K^+ ATPase (Ogawa, Shinoda, Cornelius, & Toyoshima, 2009). Although ubiquitous, cardenolides also are inducible by monarch caterpillar herbivory (Agrawal et al., 2012). Monarch caterpillars, however, together with several other milkweed specialist insects (e.g. *Rhyssomatus lineaticollis* and *Chrysochus auratus*) acquired resistance to cardenolides by at least one amino acid substitution (N122H) in ATPα (Dobler et al., 2012; Holzinger, Frick, & Wink, 1992; Zhen et al., 2012). Moreover, monarch caterpillars actively sequester cardenolides in their integument by a yet unknown mechanism (Frick & Wink, 1995; Petschenka & Agrawal, 2016; Roeske, Seiber, Brower, & Moffitt, 1976). Not only are monarchs resistant to these allelochemicals, but their sequestration of them reduces the monarch's palatability to bird predators (J. V. Z. Brower, 1958). Interestingly, adult butterflies, when they are infected by *Ophryocystis elektroscirrha*, an obligate protozoan parasite, prefer to oviposit on plants with higher cardenolide concentrations (Lefèvre et al., 2012). Infected caterpillars prefer to feed on a host with higher cardenolide concentration which translates into smaller loads of parasite spores (De Roode, Pedersen, Hunter, & Altizer, 2008).

The coevolutionary interaction between monarchs and milkweeds arose in North America almost one million years ago (Agrawal, Ali, Rasmann, & Fishbein, 2015; Zhan et al., 2014). This pair of lineages provides the best known example of plant-herbivore coevolution through the stepwise evolution of cardenolides' sequestration (Petschenka & Agrawal, 2015) linked to three discrete levels of increasingly resistant Na^+/K^+ ATPase (Petschenka et al., 2013; Petschenka & Agrawal, 2016). More than 130 species of *Asclepias* are found in North America, including Mesoamerica and the Caribbean (Blackwell, 1964; M Fishbein, Chuba,

Ellison, Mason-Gamer, & Lynch, 2011; Mark Fishbein, Juárez-jaimes, & Alvarado-, 2008; Woodson, 1954). *Asclepias* can vary more than ten-fold in their cardenolide content (Agrawal et al., 2012) and also vary in their amount of latex exudation and trichome density (Agrawal et al., 2015).

Among monarch hosts, the tropical milkweed (*A. curassavica*) was a popular choice among monarch butterfly enthusiasts because it can provide food year-round for monarch larvae (Dara A. Satterfield, Villablanca, Maerz, & Altizer, 2016). A year-round available habitat, however, can deter butterflies from migrating, resulting in resident populations with a higher proportion of *O. elektroscirrha* infection (D. A. Satterfield, Maerz, & Altizer, 2015). Paradoxically, *A. curassavica* was the host which infected monarch caterpillars prefer to feed upon (De Roode et al., 2008). Additionally, monarch larvae fed *A. curassavica* have more body mass than those fed *A. incarnata* and *A. syriaca* (Erickson, 1973; Petschenka & Agrawal, 2015), two milkweeds upon which the majority of migrants feed (Seiber et al., 1986). Another milkweed, *A. linaria*, which like *A. curassavica* is part of the American milkweed clade, was the poorest host considering caterpillar body mass (Petschenka & Agrawal, 2015) and survival to adulthood (Tao, Hoang, Hunter, de Roode, & Cotter, 2016). Regarding survival to adulthood, *A. curassavica* was only superseded, among eight milkweed species, by a host of the African clade, *Gomphocarpus physocarpus* (Tao et al., 2016). On the other hand, *A. asperula* was the worst among 53 host species, including *A. linaria*, for larval body mass after five days of feeding (Agrawal et al., 2015). In another study, however, larval body mass on *A. asperula* superseded *A. linaria* after five days of feeding (Petschenka & Agrawal, 2015). Hence, discrepancy among host rankings can exist even within the same laboratory.

Comparing transcriptomic responses of monarch caterpillars to different hosts may identify additional genes involved in cardenolides sequestration or detoxification of more general plant defenses encountered by monarch caterpillars on different milkweeds. Identifying genes underlying the monarch-milkweed interaction, would pave the way for studying their coevolution. Additionally, the genetic diversity of these genes could be subject of future studies to better inform monarch conservation efforts. To achieve this, I examined the differences in abundance of micro- and messenger RNAs when caterpillars feed

on different milkweeds for 24 hours. I selected, based upon studies described above, three milkweed species encompassing a wide range of larval growth performance and phylogenetic distances.

AIMS

General aim

Identify monarch butterfly genes underlying the performance differences observed in larvae fed on different milkweeds.

Specific aims

1. Identify protein-coding genes differentially expressed between larvae fed on selected host milkweeds.
2. Annotate microRNAs in the monarch genome.
3. Identify microRNA genes differentially expressed between larvae fed on the different hosts.
4. Identify protein-coding genes that are targets of differentially expressed microRNAs.

METHODS

Larval rearing conditions

Danaus plexippus eggs were collected from *A. curassavica* in Guanajuato, Mexico (González-De-la-Rosa, Ramírez Ramirez Loustalot-Laclette, *et al*. *manuscript in preparation*). Eggs were allowed to hatch in petri dishes. Larvae then were fed with fresh *A. curassavica* leaves until reaching the 2nd instar stage. They then were fed for 24 hours with fresh leaves of either *A. curassavica*, *A. linaria* or *G. physocarpus* (Figure 1).

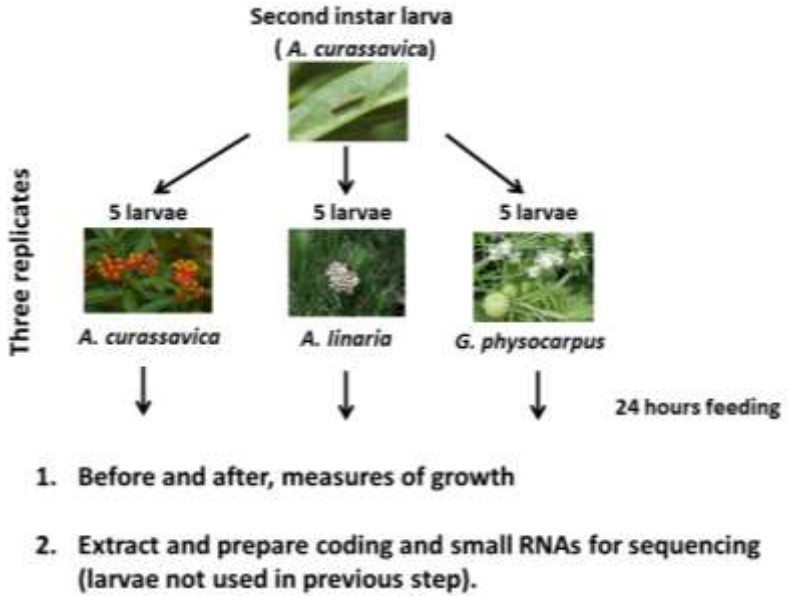

Figure 1. Overview of experimental design.

RNA extraction and sequencing

Larvae were collected after 24 hours of feeding on the different plants, immediately snap-frozen in liquid nitrogen and stored at -80ºC until the total RNA extractions were performed. Triplicates of five whole larvae for each treatment were homogenized using a TissueLyser kit (QIAGEN). Total RNA was extracted using the Direct Zol kit (Zymo). The libraries were obtained using the TruSeq RNA kit (Illumina Inc., San Diego, CA, USA) at Langebio Core Facility, and sequenced on an Illumina HiSeq™ 2500 using paired 150 nt long reads (GENEWIZ). The same total RNA was used to prepare sRNA and mRNA libraries. The small RNA

libraries were obtained using the TruSeq Small RNA kit (Illumina Inc., San Diego, CA, USA) and sequenced on an Illumina HiSeq™ 2500 using a 1x50 nt format (GENEWIZ).

Quality filtering and adapter removal

The first step for processing Illumina sequencing data is to remove residual sequence coming from the adapters used for the amplification step before actual sequencing. I assessed the adapter content and sequence quality of RNA-Seq and sRNA libraries with FastQC (Andrews, 2010). To visualize in aggregate the quality, alignment and quantification of reads I used MultiQC (Ewels, Magnusson, Lundin, & Käller, 2016). Although most , bases sequenced with Illumina are of very high fidelity, a fraction of reads will have lower quality at the 3' end due to asynchrony in the sequencing reaction. Hence, removing bases of low quality (also known as sequence trimming) reduces the likelihood of misidentifying the sequenced molecules..

The trimming strategy that I used for RNA-seq data, generally referred to as a sliding window approach, consists in assessing the mean quality of 4 contiguous bases starting from the beginning of the read. If the mean quality of the 4 bases falls below a PHRED score of 15 then the read is cut up to the last nucleotide included in the previous window where the mean quality was above 15. The PHRED score equals the logarithm base 10 of the probability of an error in the base multiplied by minus 10. Hence, given my quality threshold, all 4 base windows in the trimmed reads have an error probability below 3.16%. Here I used Trimmomatic (Bolger, Lohse, & Usadel, 2014) to remove adapters found in the file TruSeq2-PE.fa that this program provides, and to remove low quality bases as described above. Also, I discarded reads that contained less than 30 bases after adapter removal as well as those where only one pair survived after these filtering steps.

To remove bases of low confidence and trim adapters from sRNA-seq libraries, I used reaper v16-098 (Davis, van Dongen, Abreu-Goodger, Bartonicek, & Enright, 2013). To remove adapter from any part of the read, the read had to match the adapter sequence ("TGGAATTCTCGGGTGCCAAG") at least in twelve nucleotides with up to one mismatch and zero gaps (-3p-global 12/1/0/0).

However, it also trimmed the sequence if it matched eight nucleotides, with up to one mismatch, of the adapter at the start of its sequence (-3p-prefix 8/1/0). It cut no nucleotides if only the first few nucleotides of the adapter sequence were identically found at the end of the read (-3p-head-to-tail 0). Like in the case of RNA-seq, I used a sliding window to cut the sequence once the median quality of four contiguous nucleotides was below 15 (-qqq-check 15/4). Additionally, I removed reads of complexity lower than 20, as scored by DUST (-dust-suffix 20). I discarded reads with less than eleven nucleotides after quality filtering and adapter removal. The aim of removing short reads is to have more precise estimates of the fraction of reads aligning to the genome. If a big fraction of the library is composed of reads of only one nucleotide, we may incorrectly assert that most reads came from the genome because, after adjusting parameters for accepting extreme multimappers, these would align to the genome. A reasonable approach would be to choose a length threshold in which the reads were unlikely to be found just by chance given the size of the genome. I consider this arbitrary length threshold, 11, to be useful for estimating the library fraction that did not align to the genome.

Read alignment

This step consists of identifying the region in the genome that could give rise to the sequenced read. Poly(A) enriched RNA-seq libraries must be processed differently from sRNA-seq because the former includes transcripts that might have had introns spliced out, thus no longer mapping continuously to the genome. Splice-aware RNA-seq aligners can split the reads that likely span multiple exons. I used HISAT v2.1 (Pertea, Kim, Pertea, Leek, & Salzberg, 2016) to align RNA-seq reads against the soft-masked monarch butterfly genome v3 downloaded from lepbase (Challis, Kumar, Dasmahapatra, Jiggins, & Blaxter, 2016). I set the maximum intron length to 100 Kb, as recommended by Brian Haas for arthropod genomes. I allowed up to 1,657,094 (genome size divided by read length) multimapping sites to avoid multimapping reads being classified as unmapped due to being extreme multimappers. For the sRNA data, I aligned the reads with bowtie v1.2.2 (Langmead, Trapnell, Pop, & Salzberg, 2009). I allowed up to one mismatch. By being short and some being derived from repetitive regions, several short RNAs are expected to align to multiple sites. I used ShortStack v3.8.2

(ranmax = 3) to assign multimapping short reads when one of the alternative sites had a higher abundance of uniquely mapping reads nearby (Axtell, 2013).

Gene expression quantification

The approach I used to assess the expression of miRNAs was to count, with FeatureCounts v1.5.1 (Liao, Smyth, & Shi, 2014), the number of reads overlapping each mature miRNA. Furthermore, because a 3' adapter is attached to the RNA molecules prior to amplification, the result is a stranded library. Hence, I specified forward strandedness to FeatureCounts.

However, there are two approaches to quantifying expression from RNA-seq: at gene level or transcript level. Here, quantification at transcript level was unfeasible because the official gene set v2 (OGS2) annotation lacks isoforms. Gene level quantification consists of counting the number of reads per exon and then adding all the counts over all exons of each gene. I used FeatureCounts to summarize the number of reads overlapping exons of each protein-coding gene. However, for some of the paired reads taken as input by HISAT2, only one read of the pair aligned to the genome. This would happen if the pair came from a region where the assembly is incomplete or if only one of the pairs contained contaminant sequence and would imply that the sequenced RNA molecule was a hybrid between monarch RNA and a contaminant source. Because of this, I only considered reads were both pairs aligned to the reference genome.

Statistical differences of gene expression

The purpose of a differential expression analysis is to identify which genes have different expression levels between the sequenced libraries, due to the different biological conditions and not due to technical artifacts (e.g. different sequencing depth). Before interpreting differential expression results it is important to explore the relationships of the transcriptomic profiles between samples. This exploration allows us to identify if there are systematic biases (e.g. batches due to preparation of some samples at different days) by observing which samples are more similar to each other. Ideally, samples subjected to the same treatment will be highly similar yet different to samples of the other treatments. If the biological replicates (samples of the same group) are highly similar and distant from

samples of other treatments, this suggests that the most important variables were captured in the experiment and hence one can assert with high confidence that the discovered gene expression differences are due to the treatment. However, if samples of the same group are as distant to their replicates as to samples of other groups, this suggests that the effect of the controlled variable is comparable to that of factors not controlled in the experiment.

Before exploring transcriptomic profile differences, it is important to realize that we want to summarize thousands of gene expression differences in few dimensions that capture the similarity of the profiles. The most commonly used approaches are principal component analysis (PCA) and multidimensional scaling (MDS). PCA maps the data linearly to a lower dimensional space while retaining most of the variance found in the original space. One attractive feature of the PCA is that we know how much of the variance is captured by each of the new dimensions. For example, if the principal component 1 captured 100% of the variance, it would be enough to see the distance in this line between the samples to see how dissimilar, averaging across the thousands of genes, they are to one another. MDS, also known as Principal Coordinate Analysis (PCoA), is a non-linear mapping that, instead of identifying the dimensions where the samples are most distant between each other, tries to preserve the Euclidean distance between the samples in the original space but represented in N-dimensions. For example, if we have a matrix of distances between cities, the distance to one city would be one dimension of each sample. By using MDS, we could reduce the dimensions to only two and this would reflect the map where the cities lie, although disregarding cardinal directions. I considered MDS to be a better representation of the data dissimilarity as it conveys as much as possible their original distance in the full dimensional space.

I used edgeR v3.18.1 (Robinson, McCarthy, & Smyth, 2009), on R v3.4.0 (Team, 2015), to assess the statistical significance of the gene expression differences (FDR 5%). An important step before calculating the difference between treatments is to adjust the read counts to remove systematic effects (i.e. sequencing depth); this is called normalization. edgeR 1) takes the sample with closest profile to the average expression of all samples as reference; 2) calculates the log ratio (M-value) between the reference and each other sample;

3) removes genes with the most extreme expression values; 4) removes genes with the most extreme M-values; 5) estimates a normalization factor based on the weighted mean of the remaining M-values. This method is known as trimmed mean of M-values (TMM) and allows to account for the relationship between higher expressed genes having lower variance while disregarding outlier values. Prior to normalization, I filtered out genes with zero counts per million (CPM) in more than five libraries. To account for variability in library quality, measured as variability relative to a typical sample, I used voom v3.32.2 with weights (Law, Chen, Shi, & Smyth, 2014; Liu et al., 2015). To visualize the transcriptomic profiles I used the plotMDS from the limma package with “gene.selection” set to “common” to do the principal component analysis and set to “pairwise” to obtain the multidimensional scaling. In both cases I used all the genes obtained from the voomWithWeights function.

Gene set enrichment

Testing differential expression yields lists of genes that vary between the conditions of interest. When these lists include thousands of genes, it can be quite complicated to interpret these differences. One approach is having a hypothesis of which genes might be important due to previous research. A similar approach would be to skim through the list and try to come up with an explanation. Although both approaches can provide insights, they are biased to a limited subset of the results. A more comprehensive approach is to test whether genes of previously characterized pathways or processes are found in higher proportions than expected by chance. The Gene Ontology (GO) is an effort to classify the role of genes in three main categories: biological process, molecular function and cellular component (Ashburner et al., 2000). On the other hand, the Kyoto Encyclopedia of Genes and Genomes (KEGG) group genes according to cellular processes (e.g. cell cycle and metabolism) (Ogata et al., 2000). Here, I used the GO annotation available in MonarchBase and KEGG annotation from UniProt (Bateman et al., 2017; Zhan & Reppert, 2013). Furthermore, I was particularly interested in hippo, foxo, tor and apoptosis signaling pathways, but I considered that the agnostic annotation approach could be improved by identifying homologues specifically from *Drosophila melanogaster* and/or *Bombyx mori*. I downloaded from KEGG the protein sequences of these genes

and used blastp v2.2.31+ with default parameters except for “-ungapped” to identify their homologs in the monarch proteome. To further refine annotation of detoxification categories, I used the GSTs and UGTs monarch annotation recently published by Yu *et al.*, 2016. Additionally, I classified as cytochrome P450 those genes which in their name contained “CYP” or “cytochrome” or “P450”. This only misclassified one gene as cytochrome P450: “NADPH-cytochrome P450 reductase”, which I manually removed. However, because there was only one misclassified gene, this flaw is unlikely to be determinant for the enrichment results. Similarly, for alcohol dehydrogenases and carboxyl esterases, I checked that the name contained “ADH” or “alcohol dehydrogenase” or “carboxyl esterase”.

I opted for functions that test whether the observed number of differentially expressed genes (DEGs) from a given category is bigger than the number expected by chance due to the size of the category and the total number of differentially expressed genes. By enrichment of a category among up or downregulated genes, I mean that there were more DEGs of the given category than expected by chance, given the size of the given category, among the up or downregulated DEGs.

To test enrichment of KEGG pathways I used the kegga function from the limma package, while for GO I used the topGO package with the classic algorithm and Fisher’s test (Alexa & Rahnenfuhrer, 2016; Chen, Lun, & Smyth, 2016). The kegga function returns the significance of the enrichments adjusted for multiple testing. On the other hand, topGO does not adjust for multiple testing because, due to hierarchical and many-to-many relationships, the tests are not independent and hence traditional multiple-test corrections can be misleading. Furthermore, I find ambiguous whether to correct the p-values of the three gene ontologies (biological process, molecular function and cellular component) and KEGG pathways or to consider each one independently. The predominant strategy is to correct the significance of each ontology independently. I went with the default by taking the kegga results adjusted for multiple testing and leaving topGO results unadjusted. After adjusting the p-values for multiple testing, no biological process had an FDR < 5%. Therefore, I considered the topGO results

as hints of which categories were more likely to be involved. However, I rely with more confidence on the results from kegga.

Characterization of unaligned short reads

Because a considerable fraction of the sRNA-seq libraries did not align to the genome, I explored what could explain this. Because I was interested in the role of miRNAs, I checked whether some unaligned reads could be miRNAs. Tangentially, ribosomal RNA is the most abundant RNA class. I assumed that by probing the unaligned sequences against rRNA databases I could identify potential contaminant sources. I focused on unaligned reads with at least twenty nucleotides because this could be considered the lower limit of a canonical miRNA length. These I blasted with ncbi-blast+ v2.2.31 against RFAM v13.0 rRNAs and miRbase high confidence miRNAs (Kalvari et al., 2017; Kozomara & Griffiths-Jones, 2014). Before building these independent databases, I removed redundant sequences within each RNA class using sequniq from the genometools suit v1.5.8 (Gremme, Steinbiss, & Kurtz, 2013). I used the same blastn parameters as those used by the web version of BLASTn for short queries (-evalue 0.001 -word_size 7 -penalty -3 -reward 1), and accepted only one target per query ("-max_target_seqs 1"). Due the preponderant abundance of rRNAs I consider, for every read, that it originated more likely from an rRNA than a miRNA and hence if a sequence had a hit to both databases, the read was classified as rRNA.

Genome annotation

Before quantifying miRNA expression, I needed to know their locations in the genome. Although in both previous monarch genomic studies they took into account miRNAs, they only made available the mature miRNA sequences disregarding their genomic locations (Zhan, Merlin, Boore, & Reppert, 2011; Zhan & Reppert, 2013). I used MapMi v159-b32 to locate the previously identified, miRBase high confidence miRNAs as well as miRNAs identified in *Cameraria ohridella* and *Pararge aegeria* (Guerra-Assunção & Enright, 2010; Quah, Breuker, & Holland, 2015). For this particular task, MapMi aligns the mature miRNAs against the genome with bowtie v1.2.2 (Langmead et al., 2009) and then predicts the pre-miRNA through scoring the thermodynamic stability of the

secondary structure. However, in the previous monarch miRNA annotation, they used adult butterflies while we sequenced caterpillars. Hence, it is feasible that I detected a different set of miRNAs. To tackle this possibility, I analized our sRNA libraries using ShortStack v3.8.2 and miRDeep2 v2.0.0.8 because they rely on sRNA-seq data together with secondary structure stability (Axtell, 2013; Friedländer, MacKowiak, Li, Chen, & Rajewsky, 2012). For identification of precursors of conserved miRNAs and *de novo* miRNA predictions, I relied on miRDeep2 with default parameters using the trimmed filtered reads. Novel predictions were included as novel miRNAs only if they had miRDeep2 scores >100 and were consistently detected in at least five libraries. For ShortStack I also used mostly default parameters, but set minimum dicer size to 18, maximum to 24, minimum coverage to 5 reads and the pad to 10 nucleotides. Because the miRNA targets are predicted relying heavily on complementarity to nucleotides at positions 2-8 of the miRNA, to make valid predictions according to this “seed” site, I made sure that the 5’ end of each mature miRNA agreed with the transcriptional data by visual inspection and, if necessary, manual delimiting of the ranges. MiR-12 went unidentified by miRdeep2 and was also not present in the previous annotation effort (Zhan & Reppert, 2013), but through visual inspection of aligned sRNA reads and MapMi results, I annotated it in the miR-12/304/283 cluster (Figure 3). In agreement with a definition used by miRBase, I considered a miRNA cluster to be a set of miRNAs transcribed from the same strand and separated by less than 10kb (Griffiths-Jones, Saini, Van Dongen, & Enright, 2008).

However, I also annotated other non-coding RNAs to reduce the chance of misclassifying the de novo miRNA predictions. I only considered as miRNA those predictions that did not overlap with regions predicted to be rRNAs, tRNAs, repetitive elements or protein-coding exons. To identify tRNAs I used tRNAscan-SE v1.23 (Lowe and Eddy, 1996) with default parameters. For identification of 5.8S, 18S and 28S ribosomal RNAs, I used RNAmmer v1.2 (Lagesen et al., 2007). For a wider variety of non-coding RNAs, I used cmsearch from infernal v1.1.1 (Nawrocki & Eddy, 2013) with default parameters. I identified novel repeats with RepeatModeler v1.0.10. I masked and annotated these and previously reported lepidopteran repeats (queryRepeats.pl -species lepidoptera) using

RepeatMasker v4.0 (Smit, Hubley, & Green, 2015) with default parameters except for using rush mode (-qq) and asking for the result in GFF format (-gff).

Furthermore, the protein-coding genes lacked annotation of its three prime and five prime untranslated regions (UTR). To predict miRNA's targets, I needed to annotate the three prime UTR because animal miRNAs mainly regulate their targets through partial complementarity to this region (Grimson et al., 2007). To do so, I used, together with our RNA-seq data, the RNA-seq data of Zhan *et al.* (2011), which was a pool of all stages of monarch development (SRX191135). I assembled the aligned reads, described above, in a genome guided manner with StringTie (Pertea et al., 2015), without initially borrowing information from the previous annotations (OGS2). Then, I discarded transcripts that spanned more than one of the previously annotated genes because I considered these could be false chimeras. With the remaining transcripts, I ran the PASA2 pipeline (Haas et al., 2003). It associated which transcripts corresponded to previously annotated genes and added UTR annotation. However, I ran it a second time, as the guide recommends (https://github.com/PASApipeline/PASApipeline/wiki/PASA), given that the first refined models might allow a better assignment of the assembled transcripts.

miRNA target prediction

I predicted miRNA targets with TargetScan 7 (Agarwal, Bell, Nam, & Bartel, 2015). This program uses local AU content, site type (i.e. 8mer, 7mer, 7mer-1A, 6mer), three prime supplementary pairing, distance from the end of the three prime UTR, target site abundance and seed-pairing stability to assign a context-plus score. To extract the three prime UTR sequences, I first converted the GFF output of PASA2 to GTF with the convert utility of Mikado (Venturini, Caim, Kaithakottil, Mapleson, & Swarbreck, 2017). I filtered the features containing 3UTR with grep and then used the resulting GTF file with gffread, from the cufflinks suite, to extract the sequences from the monarch genome (Trapnell et al., 2010). For target prediction, I used the sequences of all the mature miRNAs, skipping conservation information and finally filtering out predictions with a context++ score ≥ -0.6.

RESULTS AND DISCUSSION

Sequencing quality and depth

For the RNA-seq libraries, I obtained between 12.8 and 16.9 million pairs of reads (Table 1). For sRNA-seq libraries I obtained between 24.5 and 35.2 million reads. All libraries had a median Phred score clearly greater than 30 across all positions (Figure 2). This means that most of the reads of all libraries had less than 0.1% probability of error in all their positions.

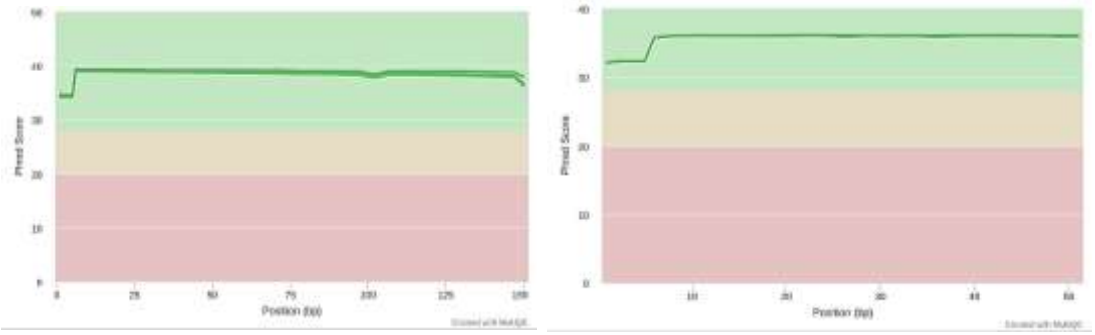

Figure 2. Median quality scores of RNA-seq (left) and sRNA-seq (right) reads. Each green line depicts the median Phred score of a sequencing library at each read position. Although most of them are indistinguishable, there are 18 green lines for the RNA-seq libraries because they are paired-end reads, while there are 9 green lines of sRNA-seq libraries. Green, beige and pink background sections are visual aids of common thresholds for good, questionable and poor quality, respectively. The Phred score equals the logarithm base 10, multiplied by −10, of the error probability (e.g. a Phred score of 30 is equal to 1 error per 1000 bases) (Ewing et al., 1998).

The adapter sequence can only be found in the read when the fragment to be sequenced is shorter than the read length of the sequencer. Because sRNA-seq typical fragments range from 18 to 30 nucleotides while the read length is fifty, I expect most of these to contain adapter sequence (Jiang, Lei, Ding, & Zhu, 2014). Most sRNA-seq reads, between 85 and 93%, contained adapter sequence. In contrast, only between 3 and 7% RNA-seq reads contained adapter sequence. (Figure 3). I have observed a similar percentage of adapter content in other sRNA- and RNA-seq datasets and therefore I consider that this does not warrant any caution about technical artifacts. The high proportion of reads with adapter

sequence in sRNA-seq reads reflect the effectiveness of size selection: most RNA fragments measured less than 50 nts. This is relevant because the short RNA molecules I want to detect, microRNAs, measure at most 24 nucleotides. Hence, miRNA sequences are likely found among sRNA-seq reads.

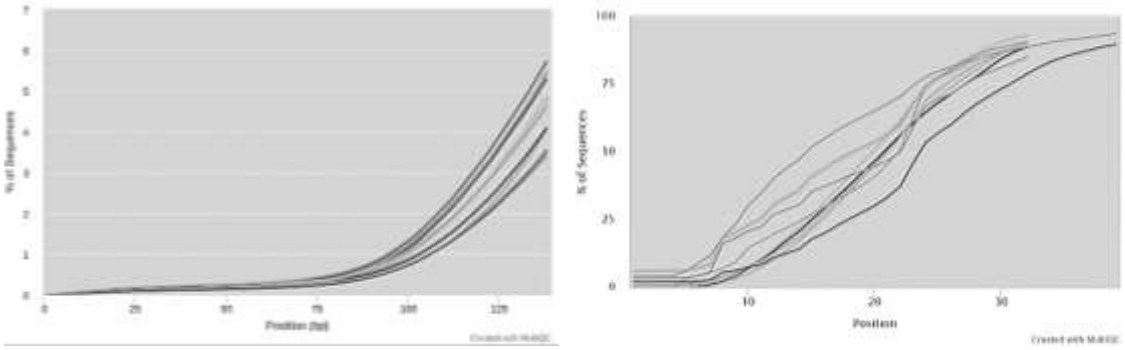

Figure 3. Adapter content in RNA-seq (left) and sRNA-seq (right) reads. Each line depicts a library as in Figure 2. The x-axis refers to the position in the read where the adapter sequence starts.

More than 95% of the RNA-seq reads survived after removing low confidence bases and adapter sequence (Table 1). However, sRNA-seq libraries AL-3, GP-1 and GP-3 lost more than 20% of their reads (Table 2). Interestingly, most short GP-1 and GP-3 extremely short reads measured seven nucleotides (Figure 4). The main sequence of this length, being more than 100 times as abundant as the second most frequent sequence, was “CCCGTGG”. Due to its brief sequence, it cannot be confidently assigned to a single genome and hence one cannot identify the organism which might have produced it. Although its extreme abundance could be considered indicative of a sequencing artifact, its absence from the other libraries makes this hypothesis unlikely. Hence, this short sequence might come from a contaminating source and I consider more appropriate to discard it, together with its variants, from downstream analyses. This observation supports the approach of discarding short sequences (< 11 nt in this case) when estimating the fraction of unaligned reads (details in Methods section).

Table 1. Statistics of the RNA-seq libraries. From left to right these are library naming, number of million paired-end reads that were sequenced, filtered, aligned and assigned to protein-coding genes.

Species	Library	Total million reads	Filtered (%)	Aligned (%)	Assigned (%)
Asclepias curassavica	**AC-1**	14.3	13.7 (95.8)	12.7 (92.7)	10.9 (85.8)
	AC-2	16.9	16.3 (96.4)	15.1 (96.4)	12.9 (85.3)
	AC-3	13.9	13.4 (96.4)	12.4 (92.5)	10.7 (86.3)
Asclepias linaria	**AL-1**	15.5	14.8 (95.5)	13.6 (91.9)	11.6 (85.3)
	AL-2	13.8	13.2 (95.7)	12.2 (92.4)	10.4 (85.2)
	AL-3	14.9	14.3 (96)	13.1 (91.6)	11.4 (87)
Gomphocarpus physocarpus	**GP-1**	14.3	13.8 (96.5)	12.7 (92)	10.5 (82.7)
	GP-2	12.8	12.4 (96.9)	11.5 (92.7)	9.8 (85.2)
	GP-3	15.3	14.8 (96.7)	13.7 (92.6)	11.6 (84.7)

Table 2. Statistics of the sRNA-Seq libraries. From left to right these are library naming, number of million reads that were sequenced, filtered, aligned and assigned to miRNA and rRNA genes.

Species	Library	Total million reads	Filtered (%)	Aligned (%)	miRNA (%)	rRNA (%)
Asclepias curassavica	**AC-1**	35.3	32.7 (92.6)	27.8 (85.2)	0.8 (2.9)	21.9 (78.8)
	AC-2	30.3	27.2 (89.8)	23.7 (87.0)	0.3 (1.3)	18.0 (75.9)
	AC-3	32.1	28.6 (89.1)	12.4 (43.6)	1.4 (11.3)	6.5 (52.4)
Asclepias linaria	**AL-1**	34	30.6 (90.0)	8.2 (27.0)	1.9 (23.2)	3.0 (36.6)
	AL-2	32.1	26.5 (83.6)	15.8 (59.8)	6.1 (38.6)	3.9 (24.7)
	AL-3	31.1	20.2 (65)	9.4 (46.6)	1.1 (11.7)	3.9 (41.5)
Gomphocarpus physocarpus	**GP-1**	24.5	18.1 (73.6)	12.5 (69.4)	3.2 (25.6)	4.5 (36)
	GP-2	28.6	26.1 (91.3)	16.3 (62.5)	3.9 (23.9)	6.6 (40.5)
	GP-3	27.9	21.5 (77.1)	16.4 (76.1)	4.9 (29.9)	5.0 (30.5)

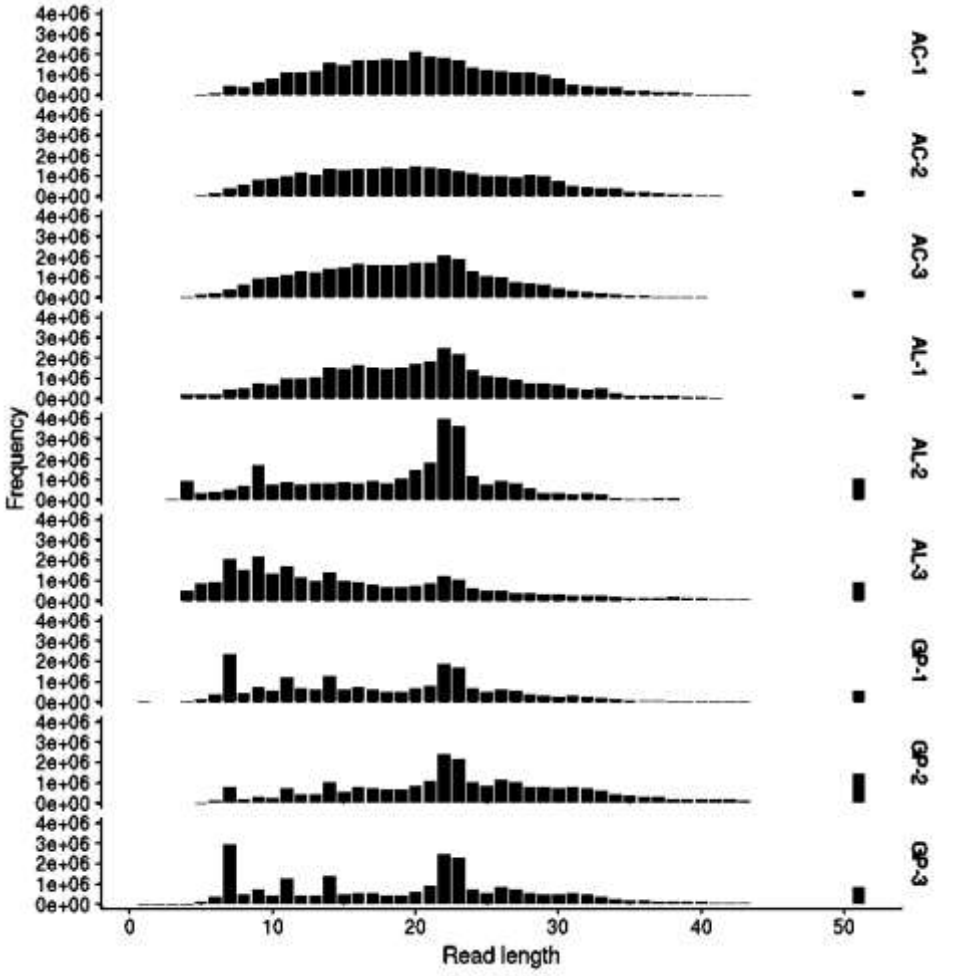

Figure 4. Trimmed sRNA-seq read length distribution.

Alignment rate

Between 92 and 96.4% of the trimmed paired-end RNA-seq reads aligned to the monarch genome (Table 1). Even in poly(A) enriched RNA-seq libraries from *Homo sapiens*, around 5% of high quality sequencing reads do not align to the genome (Granata, Sangiovanni, Thind, & Guarracino, 2017; Zhao et al., 2014). It would be more appropriate to compare my result to that of another lepidopteran or at least another insect instead of a mammal. However, although this kind of statistic is usually shown for *de novo* transcriptome assemblies (Chang et al., 2017; Smith, Macias-Muñoz, & Briscoe, 2016), it is uncommon for genome based differential expression analyses. Out of 32 articles assessing differential expression in insects, only one provided information regarding its alignment rate and it varied between 51 and 56% (Q. Y. Yu et al., 2016). Their result could be

due to sample contamination or due to genome incompleteness. Still, they do not discuss this low assignment rate at any point. Given that my result is more like the one reported by Zhao *et al.* (2014), which dealt with a model organism with high quality genomic resources, it is reasonable to assert that there is little to no contamination and that the monarch genome is fairly complete, at least regarding poly(A) RNA producing regions.

In contrast, AC-3, AL-1 and AL-3 sRNA-seq libraries had an alignment rate below 50% (Table 2). One possible explanation of the alignment rate differences between RNA- and sRNA-seq libraries is that this rate was inflated in sRNA-seq samples because the ambiguity of their short sequences. However, most of the unaligned reads have lengths clearly above (15 nts) the used length threshold (11 nts) (Figure 5). Another hypothesis is that reads were unaligned due to assembly incompleteness. Repetitive rich regions are difficult to assemble with short reads, which were used in the current assembly (Treangen & Salzberg, 2013). Given that some sRNA producing regions are found within highly repetitive sequences while protein-coding genes are generally not, it was possible that genome incompleteness would affect more severely sRNA- than RNA-seq alignment rates (Matylla-Kulinska, Tafer, Weiss, & Schroeder, 2014; Yuan, Sun, Liu, & Xie, 2011). To test this, I explored the alignment rate against an unpublished PacBio monarch draft assembly. This draft is highly redundant but, according to BUSCO (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015), is more complete (data not shown). I found that only 3.3-13.7% of the unaligned reads from each library aligned to this assembly. Hence, genome completeness cannot explain the drastically different alignment rates between sRNA- and RNA-seq libraries.

Interestingly, all the sRNA libraries from larvae fed on *A. curassavica* and two on *A. linaria* (AL-1 and AL-3) had read length distributions with a shape between uniform and a bell (Figure 4). In contrast, GP libraries and AL-2 had a peak of abundance with reads of length 22 and 23 nucleotides, the approximate size of miRNAs. AC-1 and AC-2 have almost uniform aligned read length distributions (Figure 5). Furthermore, distributions of trimmed read lengths of AC-3, AL-1 and AL-3 libraries resemble an intermediate between uniform and normal distributions. Nonetheless, when we focus only on read length distribution of the

aligned reads, these were visibly enriched in lengths 22 and 23. This highlights that only AC-1 and AC-2 lacked a clear enrichment at these lengths. One possibility is that the RNA of these libraries suffered important degradation before or during library preparation. Because rRNA is the most abundant RNA class, libraries with high degradation will be particularly abundant in rRNA fragments. Congruently, more than 50% of the aligned reads from AC-1 and AC-2 libraries mapped to rRNA (Table 2). Hence, RNA degradation could explain the lack of enrichment of reads with lengths 22 and 23 in these libraries. However, most RNA-seq reads from all libraries were mostly assigned to protein-coding genes (discussed below). Furthermore, because a considerable fraction of unaligned reads could be classified as rRNA (discussed below), it is likely that initial RNA was degraded to some extent but poly(A) enrichment filtered most of it out.

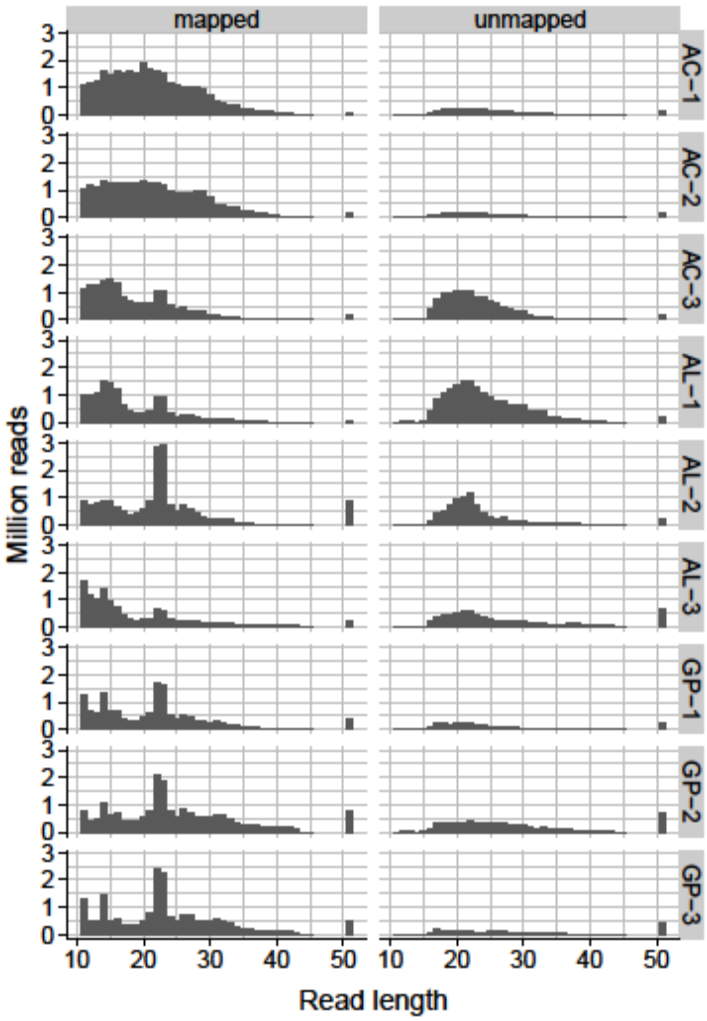

Figure 5. Mapped and unmapped read length distributions.

Another possible explanation about why these reads were unaligned was that they were a fusion of a variable RNA molecule and a fixed sequence (e.g. an adapter). If these were fused sequences, then I would find that the unidentified reads would have a length distribution like the one seen in the aligned reads but shifted by the length of the sequence to which they were fused. However, the length distribution of the unaligned reads does not resemble that of the aligned ones and is simply not shifted toward longer reads (Figure 5). Hence, most of the unaligned reads reflect fragmented non-chimeric RNA molecules.

A competing hypothesis is that the unaligned reads came from exogenous sources. To test this, I would need to identify the likely origin of the unaligned reads. This effort could be labeled as classification of metagenomic sequences and there exist specialized tools such as Kraken and Centrifuge to this end (Kim, Song, Breitwieser, & Salzberg, 2016; Wood & Salzberg, 2014). However, these tools are tailored toward longer reads. Furthermore, the results heavily depend on the database composition. In hindsight, I might have benefited from creating a database of the more likely sources of contamination. This would have included plants, particularly from the Apocynaceae family, because they could be present in the gut of larvae as there was no effort to remove the gut content. Also, I could have included known monarch parasites. Complementary to these likely contamination sources, the database should include generic contamination sources such as bacteria and fungi. However, here I opted to BLAST the unaligned reads against Rfam rRNAs and miRBase high confidence miRNAs. I chose to include miRNAs because I wanted to explore whether any potential arthropod miRNA could be found among unaligned reads. If this was the case, then those sequences could be miRNAs missing from the genome assembly. Although the most highly expressed miRNAs, the ones more likely to be detected, are the most conserved ones, metazoan miRNAs frequently show base changes outside the seed region (Liang & Li, 2009; Ninova, Ronshaugen, Griffiths-jones, & Griffiths-jones, 2014; Shen et al., 2011; Wolter et al., 2017). Hence, I could distinguish between metazoan species producing the unaligned microRNAs. Nonetheless, miRNAs are unlikely to be accurate estimators for contaminant sources due to their high expression variability even within a single specimen.

Complementary to miRNAs, rRNA is used in large-scale clinical studies of complex communities, such as human microbiota, to estimate the abundance of different taxon (Alcon-Giner et al., 2017). Hence, I took rRNA abundance as a proxy of the proportion of a contaminant species.

In several libraries, more than half of the unaligned reads could be classified as rRNAs (Figure 6). Taking as reference the rRNA fraction found among aligned reads, disregarding AC-1 and AC-2 libraries, the expected fraction of rRNA reads using this sRNA-seq protocol is between 31 and 52%. Close enough, the rRNA fraction of unaligned reads varied between 29 and 59%. This fraction is like that of the aligned reads even for the libraries with very few unaligned reads, AC-1 and AC-2. Only AL-2, the library with the clearest enrichment for reads with lengths of 22 and 23, had slightly smaller fraction of rRNA among unaligned reads. On the other hand, only a small fraction of the unaligned reads might be miRNAs. However, this small fraction is higher in AL-2 and GP libraries.

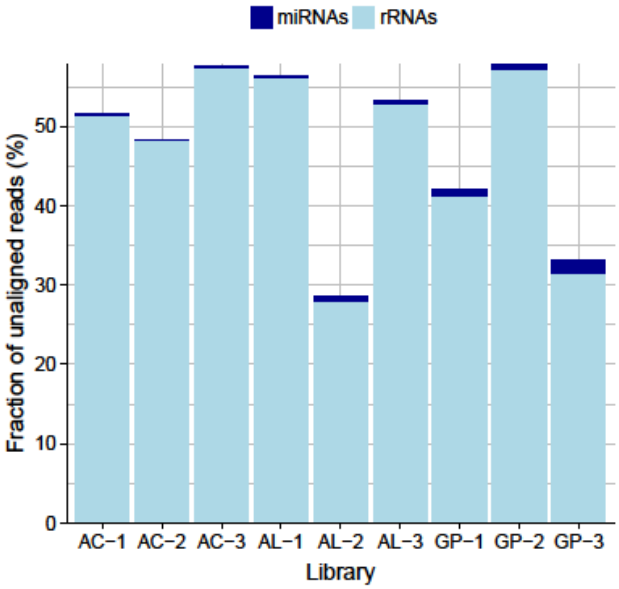

Figure 6. Fraction of unaligned reads that could be identified by blast against miRNA or rRNA databases.

Of the unaligned reads that I was able to identify as miRNAs, the most abundant sequence in GP libraries was that of miR-281 (Figure 7). Although I found this miRNA in the monarch genome, these reads did not align because they had a few non-templated adenines at the end of the sequence. This miRNA, however,

is one of the most highly expressed in all libraries. The variants of miR-281-5p were less than 0.1% as abundant as the canonical miR-281-5p. Hence, the higher fraction of miRNAs among unaligned reads in AL-2 and GP libraries only reflects their higher RNA integrity allowing to capture more miRNAs than degraded rRNAs. However, although abundant, these unaligned reads tell us nothing regarding the contamination hypotheses. In contrast, the most abundant unaligned miRNA in AC and AL libraries was a miRNA of plant origin, miR166 (Figure 7). Interestingly, this miRNA was more abundant than isomiR-281 even in the AC-1 and AC-2 libraries which had the highest alignment rate to the monarch genome. This might be interpreted, if we disregard the caveat of miRNA expression variability, either as GP libraries having less plant RNA or this miRNA having higher expression in *A. curassavica* and *A. linaria*. Assessing the contamination abundance through rRNA will favor one of these hypotheses.

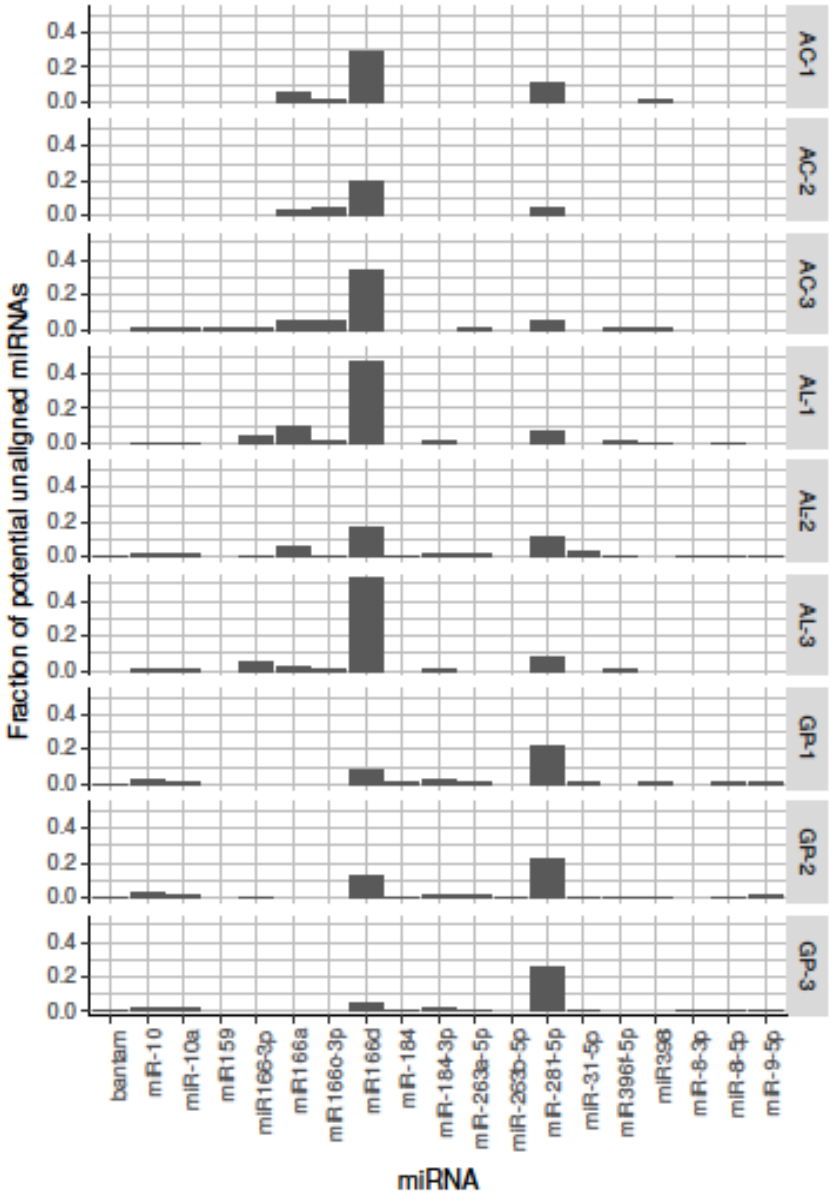

Figure 7. Most abundant unaligned miRNA reads.

In all libraries, the biggest fraction of unaligned rRNA reads came from Viridiplantae (Fig. 8A). The second most important source of contaminant RNA had no kingdom classification, but mainly consists of microorganisms (e.g. Gammaproteobacteria) (Fig 8B). The third most abundant source of rRNA was Metazoa. Finally, in all samples, a small fraction of rRNA came from fungi. However, GP libraries contained less plant rRNA and hence I consider that the difference of abundance in miR166 reflect contamination abundance rather than host-specific miRNA overexpression.

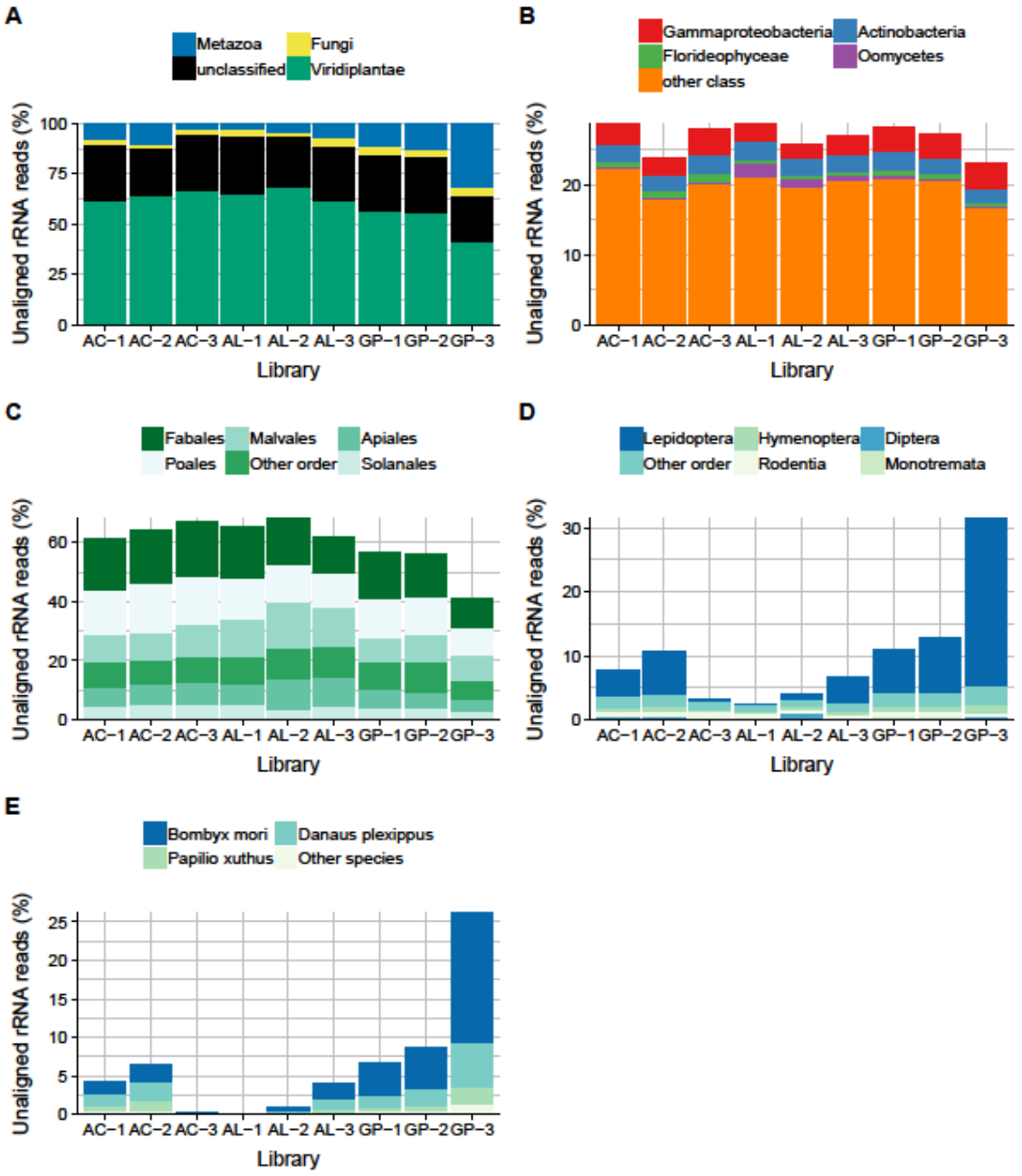

Figure 8. Sources of unaligned rRNA reads. A) Sources classified at the level of "kingdom" using taxize (Chamberlain & Szöcs, 2013), although Viridiplantae is not a kingdom in NCBI's taxonomy database. B) Most abundant orders without kingdom classification from panel A. C) Most abundant orders among Viridiplantae sources. D) Most abundant orders among metazoans. E) Most abundant lepidopteran unaligned rRNA sources.

Viridiplantae provided more than half of the identified rRNA reads in all libraries except for GP-3. Intriguingly, Gentianales order, from which all Apocynacea are

members, is not among the most abundant. After reevaluating the rRNA database I used, I noticed that it lacked sequences from the Apocynacea family. However, although the short sequences might be unable to distinguish between closely related species, they must be able to discern between different orders. Surprisingly, two very distant orders, between each other, Poales and Fabales (the latter being a monocot and the former a eudicot), were the most abundant (Figure 8C). Although this could result from contamination in the sequencer, a parameter I used for BLAST, discussed below, could better explain the apparent contamination diversity. Intriguingly, AC-3 and AL-1, the libraries with the lowest alignment rates, had the smallest fractions of Metazoa origin and virtually had no rRNA from lepidopteran origin. However, this reduction in Metazoa component is not accompanied by a clearly bigger component of plant, fungi or microorganisms. In contrast, the other libraries had a bigger fraction of metazoan origin which were mainly constituted of lepidopterans, hymenopterans and rodents. Interestingly, only the lepidopteran contribution seems to be highly variable between libraries (Figure 8D).

AL-2 stands out by having an important fraction of RNA from dipteran origin (Figure 8D). An attractive hypothesis is that this came from larvae being parasitized by tachinid flies. In particular, *Hyphantrophaga virilis* and *Lespesia archippivora* lay their eggs on foliage and can be found on laboratory reared caterpillars from eggs collected in wild plants (Oberhauser et al., 2017). Furthermore, a previous batch of monarch caterpillars reared in the laboratory from eggs collected in the same location had a high mortality rate and, in many cases, from them emerged pupae resembling those of tachinid flies. Hence, it is possible that library AL-2 included at least one caterpillar parasitized by a tachinid fly.

AC-1 and AC-2, the rRNA rich libraries with the best overall alignment to the genome, had almost the same proportion of unaligned rRNA reads coming from *Bombyx mori* as from *Danaus plexippus* (Figure 8E). In contrast, the rest of the libraries had a small fraction coming from *D. plexippus*, but most of the reads were classified as coming from *B. mori*. Given the phenotypic similarity of *D. plexippus* and *D. gilippus* and that eggs were collected where *D. gilippus* caterpillars have been observed cohabiting with *D. plexippus*, it is possible that

all libraries, but particularly GP-3, included larvae of this other species. This could be tested by assessing the alignment rate of sRNAs from a fourth or fifth instar monarch larva, when these species are easily distinguishable by the number of tentacle sets. However, I find irreconcilable that GP-3 had most of its Metazoa unaligned rRNA reads coming, supposedly, from *Bombyx mori*. I consider extremely unlikely that the samples were contaminated by a moth. This, together with the heterogeneity of plant orders detected, makes me doubt the validity of this identification strategy. A drawback of my strategy was that I did not required BLAST to align the full query, but only that the aligned segment had perfect identity. Furthermore, I only took one hit per query, although this hit is not guaranteed to be the best hit (Shah, Nute, Warnow, & Pop, 2018). Hence, I consider that the strategy I chose to explore the unaligned reads is inadequate to correctly identify their origin. Nonetheless, I consider that it pointed out two interesting hints: 1) the abundance of plant rRNA suggests that total RNA was degraded before library preparation possibly due to digested host found in the gut; and 2) AL-2 might contain at least one larva parasitized by tachinid flies. Although sample contamination should have little impact on my downstream analyses (discussed below), biological factors such as parasitism could trigger a specific transcriptomic response and hence should be considered when interpreting my differential expression results.

Still, the contamination hypothesis as explanatory of the unaligned reads leaves open the question of why this is not reflected as well in the RNA-seq. A partial explanation of why this contamination is less conspicuous in the in the RNA- than in the sRNA-seq results, is that poly(A) enrichment in RNA-seq precludes most rRNA and bacterial mRNA from being sequenced because the former are not polyadenylated while only a minor fraction of the latter harbor short poly(A) tails (Régnier & Marujo, 2013). If this was the case, then most of the unaligned reads of sRNA libraries should come have bacterial origin. The above exploration, however, suggests that only a small fraction of contamination is likely to be of bacterial origin. I also considered that high degradation of rRNA could exponentially inflate the abundance of exogenous RNA. This would only affect sRNA-seq because the protocol only selects the molecules by size while RNA-seq involves the poly(A) enrichment step. However, poly(A) RNA-seq would be

unable to detect the contaminant RNA if it was highly degraded while sRNA-seq would definitely be able to detect it. In summary, even though sRNA- and poly(A) RNA-seq libraries were prepared from the same RNA, one shows important fraction of exogenous RNA while the other does not. Most likely, this difference arise by exogenous RNA present in the gut being degraded during digestion. The size enrichment step would capture many of these degraded fragments while poly(A) selected libraries would barely detect them.

The monarch genome contains 99 microRNAs

In total, I identified 99 miRNAs in the monarch genome. Of these, 87 are conserved in other lepidopteran species. Recent studies have identified 81, 86 and 90 conserved miRNAs in *Cameraria ohridella* (moth), *Blatella germanica* (cockroach) and *Pararge aegeria* (butterfly), respectively (Quah, Hui, & Holland, 2015; Ylla, Fromm, Piulachs, & Belles, 2016). In contrast to the previous annotation, I identified six additional miRNAs conserved in lepidopterans including miR-12 from the miR-12/304/283 cluster. Furthermore, I could only reliably annotate 3 out of the 35 novel predictions from the previous annotation effort (Zhan et al., 2011). I only predict nine novel miRNA loci and hence a total of twelve monarch-specific miRNAs (Supplementary Table 1). I favored stringency over sensitivity because a large fraction of miRNA annotations are false positives (Chiang et al., 2010). However, when I noticed that miRDeep2 was unable to identify miR-12, I also realized that some conserved miRNAs identified by MapMi did not perfectly agreed with the stacked alignments (Figure 9). Before noticing this, I had already predicted targets for these mature miRNAs and found that the “seed” site of miR-278 did not agree with the one previously reported. This pointed out the importance of well-defined miRNA start and end positions. Therefore, I curated all mature miRNA ranges by visual inspection of the annotation and its agreement with sRNA-seq reads alignments. An important fraction of miRNAs, conserved and non-conserved, has been previously found within introns of protein coding genes (Chiang et al., 2010). Here I found 35 such miRNAs. Furthermore, several miRNAs exist as clusters spaced only by a few nucleotides (Lagos-Quintana, Rauhut, Lendeckel, & Tuschl, 2001; Lau, Lim, Weinstein, & Bartel, 2001; Y. Lee, Jeon, Lee, Kim, & Kim, 2002). Accordingly, I found 38 miRNAs grouped in 16 clusters, of which the highly conserved miR-

71/2a-1/13a/13b/2a-2/2b cluster was the largest one (Marco, Hooks, & Griffiths-Jones, 2012). Interestingly, although conserved miRNAs were identified independently of gene expression, all of them, except for miR-932, were detected by at least one count per million.

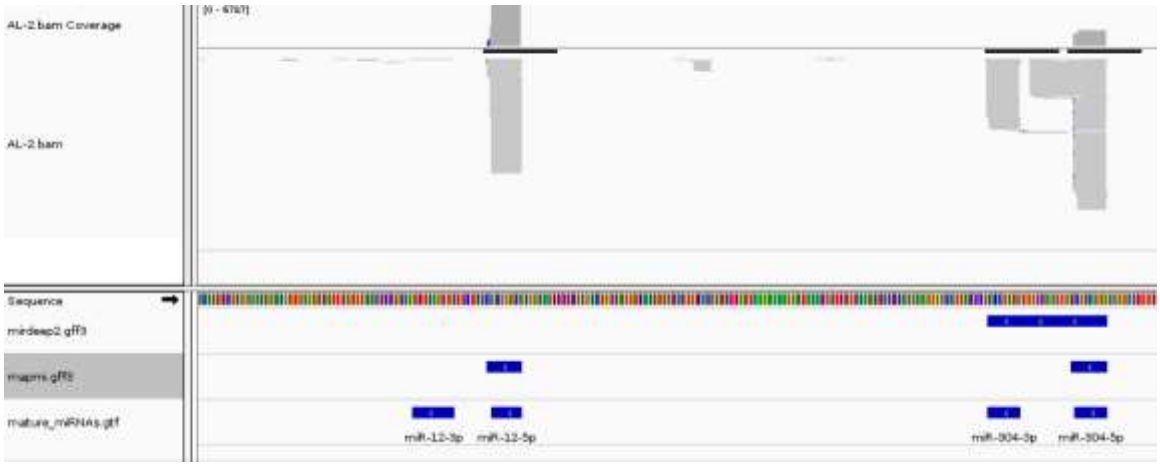

Figure 9. Neighborhood of miR-12. Screenshot of Integrative Genomic Viewer portraying the sRNA-seq the upper panel; miRDeep2, MapMi and final mature miRNA annotation are found in this order below the genome sequence track.

Protein-coding genes detected through sequencing

Of the aligned RNA-seq paired-end reads, between 82.7 and 87% were unambiguously assigned to annotated exons of protein-coding genes (Table 1). This is slightly higher than the 80% rate found by Zhao et al. (2014) in the case of humans. This suggests that the poly(A) enrichment and protein-coding annotation are adequate for our gene expression analysis.

Out of the 15,128 annotated genes in the monarch genome, 11,259 (74.4%) had more than one count per million (CPM) in at least three libraries. If my filtering criterion was having at least one fragment per kilobase per million (FPKM) in at least three libraries, then 11,063 (73.1%) genes would be selected. In contrast, Yu et al. (2016) found 7,993 (49.3%) genes with more than one FPKM in *Heliconius melpomene* larval gut. However, the libraries I analyze come from whole larvae. Given the more comprehensive mix of different cell types found in the whole larvae compared to cell types found only in the gut, I consider comparing the number of detected genes in one tissue versus whole body to not

be very useful. Similarly to the alignment percentages mentioned above, I could not find the numbers of detected genes in any of 27 RNA-seq differential expression studies using RNA-seq in insects. In order to compare my results to whole body sequencing of another insect, I searched for poly(A) RNA-seq datasets of insects in the Sequence Read Archive (SRA). I found a collection of summarized gene expression from RNA-seq libraries from eight *Drosophila* species (GSE99574). I focused only on the results of *D. melanogaster* and *D. willistoni*. I found that 13,100 (75%) and 11,739 (79%) genes of *D. melanogaster* and *D. willistoni*, respectively, had more than one CPM in at least three samples. These libraries came from whole bodies of sexually-mature freely-mated adult flies (4 to 14 days post-eclosion). Although larvae are quite different from adults, I consider a valid proxy to explore the expression in mature flies because both larvae and mature stages are composed of a wide variety of cell types in different proportions. Furthermore, actin and several ribosomal protein genes, commonly considered as housekeeping (Singh et al., 2018), were consistently found among the genes with highest normalized expression (fragments per kilobase per million) in all libraries. Hence, the number of detected protein-coding genes and their expression correspond reasonably well to what I could expect for whole body sequencing of an insect.

MicroRNA genes detected through sequencing

In contrast to protein-coding genes, only between 1.3 and 38.6% of aligned reads from sRNA-seq were assigned to miRNAs (Table 2). The libraries with lowest miRNA fractions were AC-1 and AC-2 (Table 2). Still, the low miRNA fraction of AL-2 and GP libraries is surprising because they had a clear peak of aligned reads with the known size of miRNAs (Table 2). This apparently undermines the deduction that these libraries were enriched in miRNAs or that some miRNAs remain unannotated as such. However, by focusing on the aligned reads in the range of miRNA sizes, we can see that most of the AL-2 and GP libraries were indeed miRNAs (Figure 10).

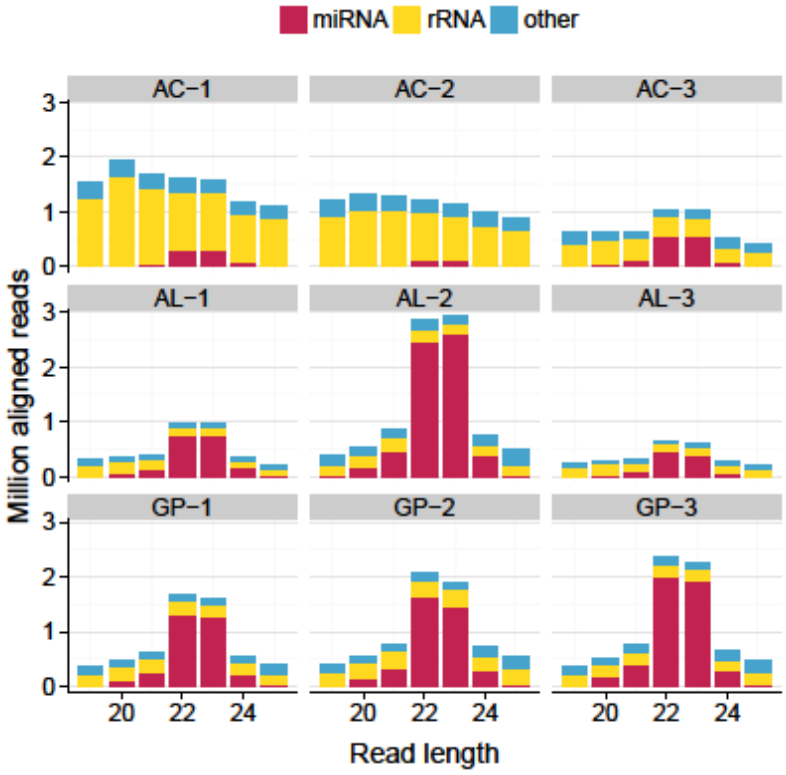

Figure 10. microRNA and rRNA biotype frequency among reads measuring between 19 and 25.

Examining the expression levels of mature miRNAs, I found that the most highly expressed miRNAs were miR-281 and miR-31 with close to 20% and 14% of total miRNA reads, respectively, followed by miR-6094, miR-10, miR-8, miR-263a and bantam (Figure 11). Particularly, bantam, miR-10, miR-31, miR-8, miR-263 and miR-281 have been previously found among the most expressed miRNAs in second instar *P. xylostella* larvae and across different stages of *Heliconius melpomene* (Etebari, Hussain, & Asgari, 2013; Surridge et al., 2011). Finding that miRNAs have relatively similar proportions in the three diets, regardless of some of them having different alignment and miRNA assignment rates, bolsters the notion (discussed below) that RNA degradation and possibly other factors (e.g. sample contamination), although decreasing the total number of miRNA reads by increasing rRNA and mRNA reads, does not alter their relative abundance.

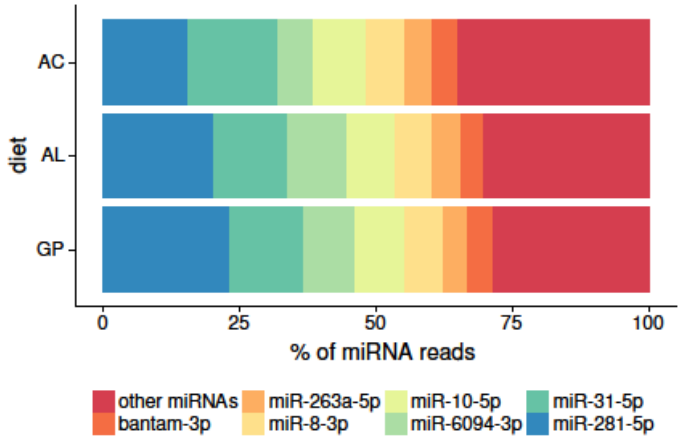

Figure 11. Most abundant microRNAs. Each diet is represented by the average of its three replicates.

It has been pointed out that extraction, sequencing and normalization of sRNA-seq data can bias differential expression results (Anfossi, Babayan, Pantel, & Calin, 2018). This was of particular concern for my results because some of the libraries had an important fraction of reads that did not align to the genome and other libraries, with high alignment rate, contained relatively few miRNA reads. Nonetheless, I assume that the smaller number of miRNA reads obtained in the degraded samples, is still a random sample of the miRNAs that we would have seen if size enrichment was effective and all contaminant RNA was absent. Finding that all libraries shared their top expressed miRNAs is reassuring in the context of this assumption (Figure 11). Given that edgeR normalizes observed counts by library size, this will account for the random sampling due to technical artifacts when assessing differential expression, if library size is calculated as the total number of reads mapping only to microRNAs. However, genes with low expression are more susceptible to variations in depth of sequencing. EdgeR also considers this when estimating the significance of the difference between treatments. Therefore, I consider that the downstream differential expression analysis can score a sensible ranking of the confidence of the miRNA expression differences between the different diets.

Furthermore, I found that several miRNAs had predominant expression of an arm which is not reported as the most abundant in miRBase for its *D. melanogaster* homolog. At some stage of miRNA research this was an interesting finding because originally it was thought that only one arm of the miRNA bound to the

RNA-induced silencing complex (Han et al., 2006; Khvorova, Reynolds, & Jayasena, 2003; Schwarz et al., 2003). Afterward it was later found that both miRNA arms can be robustly detected at least in mammals (Bhayani, Calin, & Lai, 2012; Ro, Park, Young, Sanders, & Yan, 2007). This flexibility has also been found in several insects including *P. xylostella*, *B. mori*, *Manduca sexta*, *L. migratoria* and *B. germanica* (Cristino, Tanaka, Rubio, Piulachs, & Belles, 2011; Etebari et al., 2013; Li et al., 2014; Wei, Chen, Yang, Ma, & Kang, 2009; Zhang et al., 2015). Hence, this finding agrees with previous reports and it is relevant to keep in mind whenever considering miRNA targets as these predictions tend to be performed using only one of the arms.

Transcriptomic profiles correlation

I found that when using the expression of all genes, instead of the default 500 most variable, MDS and PCA reveal the same pattern (Figure 12).

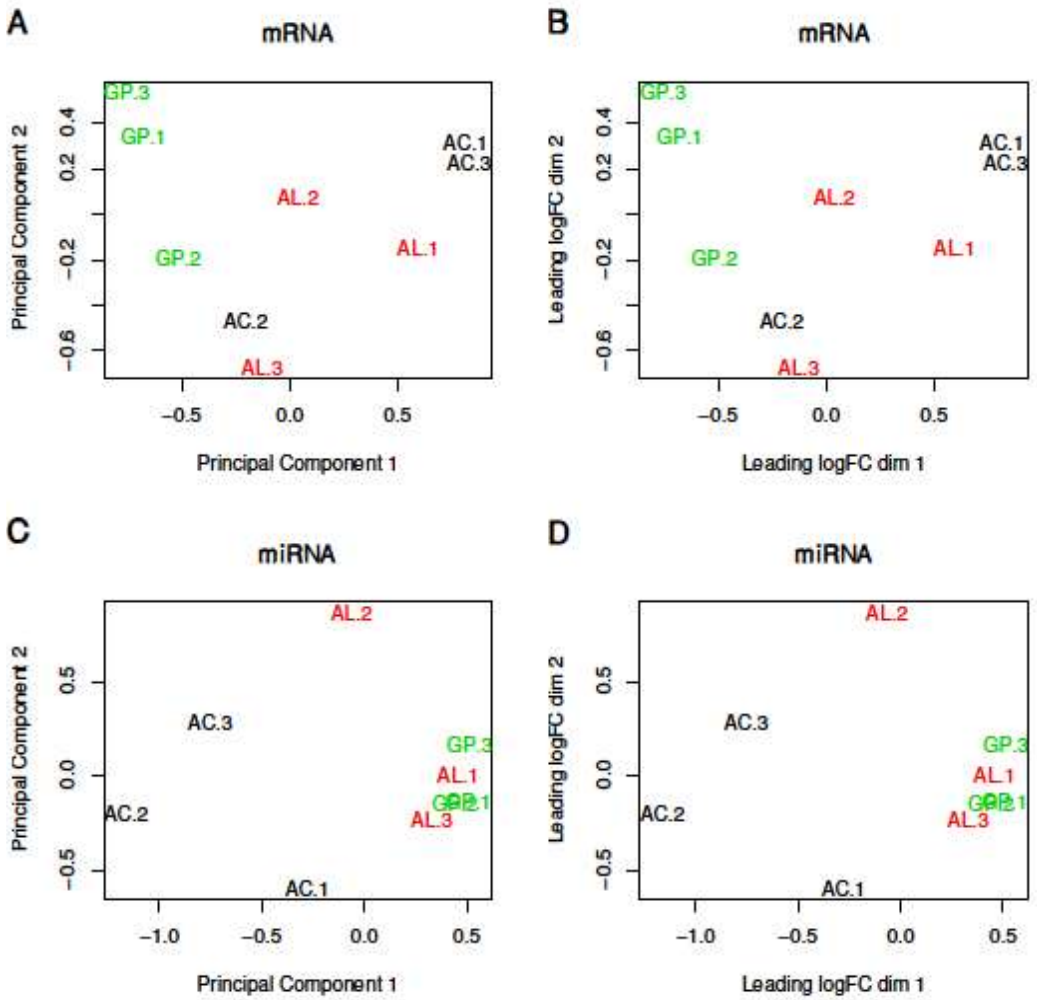

Figure 12. Sample's transcriptomic profile correlation. A) PCA of protein-coding genes' expression. B) MDS of the pairwise fold-changes. C) PCA of miRNA genes' expression. D) MDS of miRNA genes' expression. Samples are colored according to diet.

Given that miRNA and mRNA expression were assessed from the same RNA pools and that both contaminant and rRNA were excluded from quantification, I expected to find similar distances between the samples and, particularly, according to their treatment. I found that, in both approaches, *G. physocarpus* yields the most distant transcriptome relative to *A. curassavica* diet, while *A. linaria* is between them. However, in the case of miRNAs, the *A. linaria* diet resulted in profiles more similar to those of larvae fed on *G. physocarpus* instead of those fed on *A. curassavica*.

Additionally, due to the hint of having at least one parasitized larva in the AL-2 pool, I was curious about the dissimilarity of this profile within the replicates of the *A. linaria* diet. Interestingly, this sRNA-seq library is indeed the most distant to all other samples in the second principal component. In contrast, the RNA-seq profiles do not single out AL-2 in the two first principal components . However, when assessing the third and fourth principal components of protein-coding profiles, AL-2 is one of the three outliers (Figure 13). Therefore, the hint of parasitism is partially supported by miRNA (PC2) and the mRNA (PC3 and PC4) expression profiles.

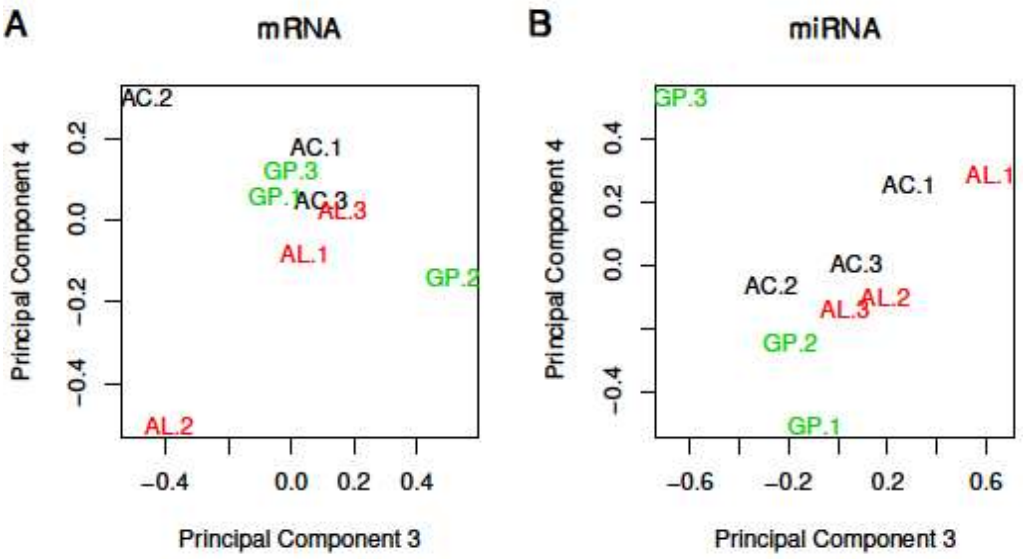

Figure 13. Sample's transcriptomic profile correlation in the third and fourth principal components. A) Protein-coding genes' expression. B) miRNA genes' expression.

Still, AC-2 was found closer to AL and GP samples than to the other AC samples as measured by expression of protein-coding genes, while it is the most distant from non-AC samples according to miRNA expression profiles. This portrays AC-2 as a sample which might be object of an unaccounted variable, at least through the RNA-seq approach. One possibility I considered was that AC-2 library might be mislabeled. Also, given that AL-1 is closer to the non-suspicious AC libraries, I also considered AL-1 as an outlier that might be mislabeled. Including mislabeled libraries would diminish the statistical power to detect the different effects of each diet. One could argue that discarding these outliers would yield a more reliable result. On the other hand, discarding these outliers can also be criticized as biasing the results by using only the subjectively cleaner samples. Also, an undesirable consequence of discarding samples is that it reduces the statistical power by reducing the sample size of each treatment. I considered this and the possibility that, instead of mislabeled larvae, these samples had different profiles due to coming from different collection sites. However, I had no evidence for relabeling the samples or to account for another factor than diet.

A good compromise between removing noisy samples and including them is to penalize the most variable samples. This approach, proposed by Liu *et al.*, (2015), shows that down-weighting the most variable samples is a better strategy than removing the outliers. I found that AC-2, AL-3 and GP-2 were the most downweighed in the protein-coding expression profiles (Figure 13A). However, as expected due to the dissimilarity of protein-coding and miRNA gene expression profiles (Figure 12A, 12C), there is little correspondence with the weights assigned to the miRNA profiles where only AC-1 and AL-2 were heavily downweighed (Figure 13B). This strengthens the notion that a technical factor that increased variability (e.g. RNA degradation) differently affected RNA- and sRNA-seq libraries. I ignore which technical factors might have impacted AC-2, but the PCA suggests that it was subject to a factor other than treatment and is downweighed in my mRNA differential expression analysis. In the case of miRNAs, I expected AC-1 and AC-2 to be the most downweighed libraries

because RNA degradation increases sample variability. These two samples indeed had smaller weights than AC-3, but AC-2 was not as penalized as AC-1 even though they have similar fractions of rRNA. On the other hand, AL-2, the sample which might contain parasitized larvae, was even more downweighed than AC-1. Hence, I consider that for my case the most suitable strategy is weighing libraries according to their dissimilarity relative to the average library of each treatment.

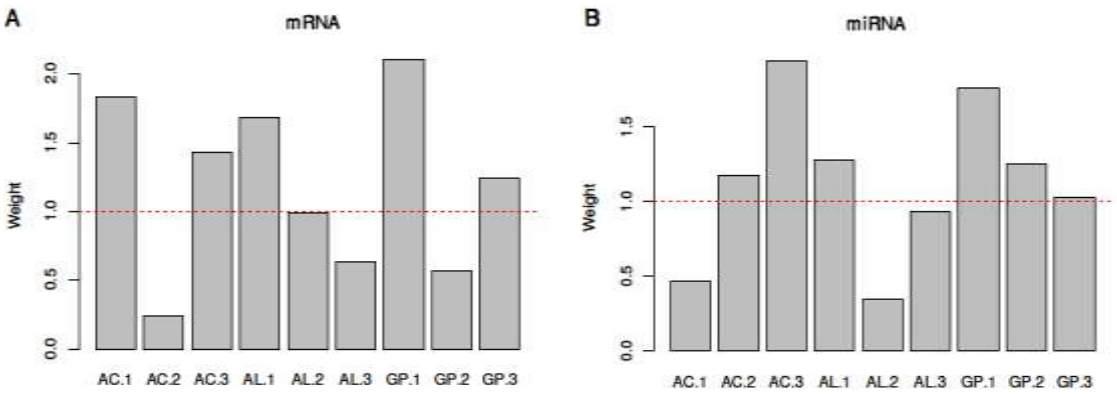

Figure 14. Sample weights of mRNA (A) and miRNA libraries (B) reflecting the variability of each sample relative to a typical sample.

Nonetheless, I explored how robust were these results to the exclusion of outliers. I found the weighting approach identified almost all the genes classified as differentially expressed when removing the outliers in both contrasts involving the *G*. *physocarpus* diet, and further detected an important additional number of differentially expressed genes (Table 3). Only the AL vs AC contrast showed the opposite trend, where excluding outliers detected more differentially expressed genes than using weights. Hence the result of the AL vs AC comparison is severely affected by the decision of excluding or including outliers. Still, it is understandable that both approaches have a much smaller intersection in this contrast because these treatments were the ones which included both outliers.

Table 3. Number of differentially expressed genes excluding outliers or using the weighting approach. FDR < 0.05.

Contrast	Direction	Intersection [a]	Weighting [b]	Outliers removal [c]
AL vs AC	Up	35	22	38

	Down	4	5	127
GP vs AC	Up	2038	924	15
	Down	2271	578	176
GP vs AL	Up	789	1370	28
	Down	1332	1040	20

[a] Number of DEGs identified through both approaches. [b] Number of DEGs identified exclusively through the weighting approach. [c] Number of DEGs identified exclusively when removing outliers.

Gene differential expression

Of the 66 protein-coding genes differentially expressed in the AL vs AC contrast, only nine reduced their expression in larvae fed on *A. linaria* (Supplementary Table 2). The annotated genes with most significant differential expression included E3 ubiquitin-protein ligase, Bloated tubule and Alanine-glyoxylate aminotransferase 2 like protein. The first might be related to protein degradation or cell cycle progression or transcriptional regulation or DNA repair or signal transduction; the second is a neurotransmitter with sodium symporter activity and the last one is related to nicotine resistance (Morreale & Walden, 2016; Passador-Gurgel, Hsieh, Hunt, Deighton, & Gibson, 2007; Thimgan, 2006). In the opposite direction, several of the annotated genes are related to carbohydrates and fatty acid metabolism, suggesting that larvae obtain different nutrients and possibly different allelochemicals from the different *Asclepias*.

I applied this same strategy to assess the differential expression of miRNAs. Comparing GP vs AC identified 49 DE miRNAs (Supplementary Table 3). Twenty-three of these have higher abundance in larvae fed GP. Similar to the RNA-seq result, the contrast between *Asclepias* identified very few differentially expressed miRNAs: only miR-novel36-5p, miR-novel28-3p and miR-6094-3p. However, unlike the RNA-seq result, comparing GP vs AL only identified miR-278-5p and miR-novel37-5p, as being down and upregulated, respectively. This resonates with the differences observed in the PCAs: AL diet had a similar effect to GP when assessed through miRNA expression profiles but differed drastically when measured through expression profiles of protein-coding genes (Figure

12A). However, having the RNA-seq result as background, I expected a disproportionally higher number of differentially expressed genes in both contrasts involving GP diet while having very few differentially expressed between the *Asclepias*. Nonetheless, if the effect of the *A. linaria* diet over the larvae's miRNA expression profile was so dissimilar to that of *A. curassavica*, as suggested by the PCA, and like that of *G. physocarpus* diet, then I would expect similar numbers of differentially expressed miRNAs in both contrasts involving *A. curassavica*. Furthermore, I would expect many of these differentially expressed miRNAs to be identified in both contrasts. Neither of these expectations were met. The poor overlap between the miRNAs identified by the contrasts involving AC suggests that AL and GP indeed are dissimilar conditions. If the effect of AC diet indeed was different to AL, at least at the level of miRNA gene expression, then the AC vs AL contrast would identify the dozens of miRNAs that are not found in common in the GP vs AC and GP vs AL. However, this was not the case and hence I doubt the robustness of most miRNAs classified as differentially expressed in the GP vs AC contrast. Given that most of the differentially expressed protein-coding genes were identified in both contrasts involving GP, that this result was robust to outliers' inclusion or exclusion, and that this approach was mostly devoid of exogenous RNA, I trust more in the miRNAs identified by both contrasts than only by GP vs AC, which include many more. Hereon, I only trust as differentially expressed those miRNAs that were either identified by both contrasts involving GP or those that were differentially expressed between AC and AL and also identified in the GP vs AC contrast (Figure 14). With this conservative approach, only miR-278-5p and miR-novel37-5p were identified as potentially involved in regulating some of the thousands of protein-coding genes differentially expressed between *Asclepias* and *G. physocarpus*. On the contrary, miR-novel36-5p, miR-novel28-3p could regulate some of the hundreds of protein-coding genes that differed between larvae fed on the distinct species of *Asclepias*.

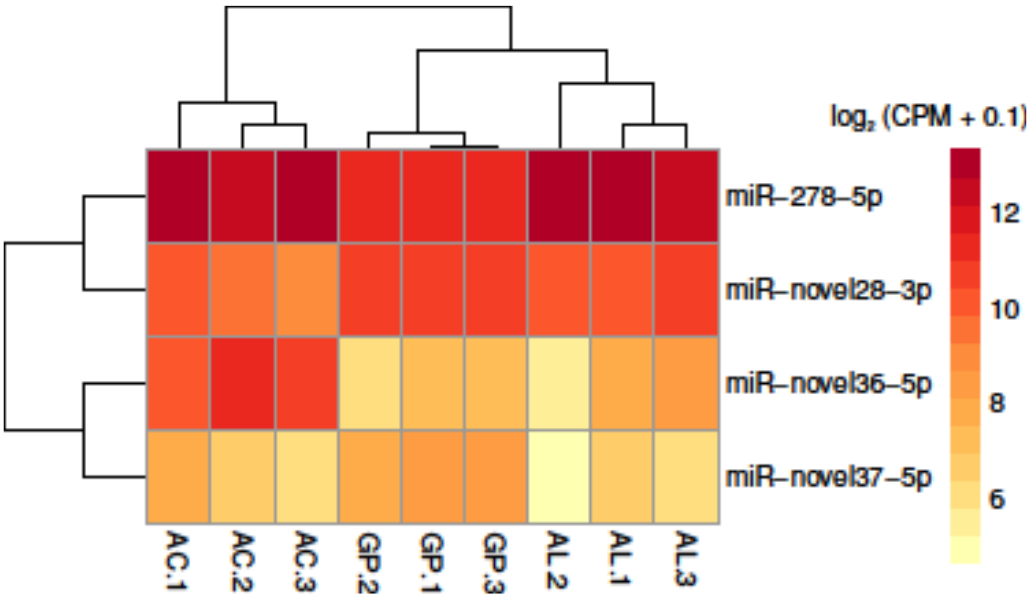

Figure 15. Expression profile of differentially expressed miRNAs. FDR < 0.05. CPM: counts per million.

Growth related genes were downregulated in larvae fed on *Gomphocarpus physocarpus*

In order to interpret the biological relevance of the differentially expressed genes, I explored which pathways and gene ontologies were enriched amongst them. I performed enrichment tests of the Biological Processes Gene Ontology (GO) and KEGG pathways, assigned by MonarchBase and UniprotKB, respectively. I identified 19 KEGG pathways enriched when comparing GP vs AL and 16 when comparing GP vs AC contrasts. Thirteen of these were common to both contrasts (Table 4). Although nine pathways were highlighted in only one of these contrasts, I will not consider them downstream because they were not identified in the contrast comparing both *Asclepias*. Hence there is no robust evidence that these pathways are specific to either of the *Asclepias* diets. Only the valine, leucine and isoleucine degradation pathway was enriched among the genes upregulated in larvae fed on *A. linaria* relative to the ones fed on *A. curassavica*.

Table 4. KEGG pathways found in the enrichment analyses of GP vs AC and GP vs AL contrasts. FDR < 0.05.

Pathway	N	Up	Down	Direction	FDR
DNA replication	41	1	38	Down	$5.2x10^{-16}$
Protein processing in endoplasmic reticulum	154	18	88	Down	$1.9x10^{-12}$
Proteasome	46	1	37	Down	$3.1x10^{-11}$
Spliceosome	162	22	87	Down	$3.4x10^{-10}$
Aminoacyl-tRNA biosynthesis	53	3	36	Down	$2.8x10^{-7}$
Cell cycle	99	12	56	Down	$2.9x10^{-7}$
Nucleotide excision repair	49	3	32	Down	$1.0x10^{-5}$
Mismatch repair	23	1	19	Down	$1.4x10^{-5}$
Protein export	23	0	18	Down	$1.5x10^{-4}$
Gastric acid secretion	55	29	9	Up	$1.1x10^{-2}$
Salivary secretion	71	35	12	Up	$1.3x10^{-2}$
ABC transporters	53	28	3	Up	$1.5x10^{-2}$
Phototransduction	30	18	5	Up	$3.5x10^{-2}$

A highly relevant category, given that larvae fed on *G. physocarpus* barely grew, was *Cell cycle*. In agreement with poor growth, cell cycle genes of larvae fed on *G. physocarpus* were mostly downregulated. The four canonical cyclins involved in regulating cell cycle progression: A, B, D and E, were downregulated (Supplementary Table 4). Furthermore, an important pathway for growth regulation through restraining cell proliferation is Hippo signaling (Harvey, Pfleger, & Hariharan, 2003). Inactivation of Hippo kinase results in translocation of Yorkie to the nucleus where it transcribes genes involved in cell proliferation (Meng, Moroishi, & Guan, 2016). The decreased expression of hippo and increased expression of yorkie in larvae fed on GP (Supplementary Table 2) is consistent with the idea that larvae fed on GP spend fewer resources on synthetizing transcripts for *cell cycle*. Other enriched pathways like *DNA replication, nucleotide excision repair* and *mismatch repair* fit well with the perspective of poor cellular growth as DNA replication is essential for cell division and at the same time demands the machinery to repair arising errors (Fersht & Knill-Jones, 1981). Furthermore, the spliceosome pathway was enriched among genes downregulated in GP. Depletion of most components of the spliceosome favor early cell cycle arrest in G1 (Karamysheva, Díaz-Martínez, Warrington, &

Yu, 2015). Together, these categories portray a consistent image of decelerated cell growth.

A hypothesis explaining the limited growth of larvae fed on *G. physocarpus* is that this diet yielded insufficient nutrients. This could be consequence of more abundant or more toxic allelochemicals or due to fewer nutrients available in the host. In *Bombyx mori*, 3 hours of starvation were enough to induce an hyperglycemic response (Satake, Kawabe, & Mizoguchi, 2000). Hence, if larvae fed on *G. physocarpus* were nutritionally restricted, then genes of insulin signaling or lipid metabolism would be deregulated (Chatterjee et al., 2014; Puig & Tjian, 2006). A critical gene in this venue is Forkhead box sub-group O (foxO) because, as part of insulin signaling, it links nutrient availability with cell growth (Tang, Smith-Caldas, Driscoll, Salhadar, & Shingleton, 2011). The phosphorylation of FoxO by Akt1 results in its translocation from the nucleus to the cytoplasm where it becomes unable to regulate the transcription of its targets (Biggs, Meisenhelder, Hunter, Cavenee, & Arden, 1999; Brunet et al., 1999; Kops & Burgering, 1999). Under nutrient restriction, FoxO is found in the nucleus and there it promotes the transcription of eukaryotic translation Initiation Factor 4E Binding Protein (eIF4E-BP also known as thor), insulin receptor (inR) (Puig, Marr, Ruhf, & Tjian, 2003) and InR substrate (chico) (Marr, D'Alessio, Puig, & Tjian, 2007). Coherently, I found foxO, thor, inR and chico upregulated in the GP diet. Conversely, eukaryotic initiation factor 4E, the target of thor, was downregulated. Given that the change in expression of thor and its target suggest that fewer mRNA are being translated, it is reasonable, from a framework of optimizing resources, that fewer proteins will be degraded given that they are not being synthetized. Accordingly, almost all proteasome genes were downregulated in larvae fed on GP. Furthermore, inhibition of the proteasome causes cell growth arrest (Yin et al., 2005). Hence, the transcriptomic response of larvae fed on *G. physocarpus* acquiesce in the hypothesis that the shift to this host resulted in nutritional restriction.

Of the five miRNAs I trust as differentially expressed, only miR-278 has been previously characterized. Overexpression of miR-278-3p in *D. melanogaster* resulted in eye and wing overgrowth (Nairz et al., 2006; Teleman & Cohen, 2006). Furthermore, this miRNA links nutrient availability with cell division in germline

stem cells as part of the insulin signaling pathway (J.-Y. Yu et al., 2009). miR-278-3p was downregulated in larvae fed on *G. physocarpus* and I would expect, due to its previously reported roles, that this accompanied a reduction of larvae growth due to nutrient scarcity.

Given that the phenotypes of miRNA alterations are explained through their downstream targets, I further explored the resemblance of miR-278-3p and hypothesized the biological relevance of the non-characterized miRNAs. I identified the untranslated regions (UTRs) of the protein-coding genes using both our RNA-Seq and previously published data (SRX191135) and used TargetScan to identify and score targets, according to context features such as distance from end of the open reading frame and number of complementary sites in the 3' UTRs (Agarwal et al., 2015). I only considered predicted targets with an opposite fold change to that of its regulating miRNA (Supplementary Table 5). The underlying assumption is that if miRNA expression had a detectable difference at the body-wide level, then its effect on a target should also be detectable at the same level. This, however might not be the case because several miRNAs have tissue-specific expression and their targets are downregulated in these same tissues but can have higher expression in tissues were the miRNA is absent (Sood, Krek, Zavolan, Macino, & Rajewsky, 2006). If a miRNA is highly expressed only in a few cells, the RNA-seq approach might be unable to discern the different abundance of its targets. Nonetheless, if the miRNA was so localized, it would be equally difficult that the sRNA-seq identify it as differentially expressed. Hence I consider that the strategy of focusing on miRNA:target interactions of opposite fold-change is, at least for miRNAs that are expressed in most cells of the larva, a suitable approach to explore the function of miRNAs.

Previous reports show that *expanded*, part of the hippo signaling pathway which is critical for growth regulation, and *dacapo*, a cyclin-dependent kinase inhibitor involved in cellular division, are critical targets of miR-278-3p, and partially explain its role in altering organ size in *D. melanogaster* (Teleman & Cohen, 2006; J.-Y. Yu et al., 2009). There are two complementary sites to miR-278-3p in the 3' UTR of *expanded* gene in *D. melanogaster* (Teleman & Cohen, 2006). However, in the monarch genome I identified only one complementary site within a less favorable context for the pairing between miR-278-3p and *expanded*'s 3' UTR.

This suggests that miR-278 has a weaker effect on the regulation of *expanded* than the one reported in *D. melanogaster*, but leaves open the possibility that, in the monarch butterfly, miR-278 regulates organ size through its target *expanded*. TargetScan did not identify a complementary site to miR-278-3p in the 1392 bases of *Dacapo* 3' UTR. It is unlikely that miR-278 regulates organ size in *D. plexippus* through downregulation of *dacapo*. Instead, another gene of the Hippo signaling pathway, Kibra, was identified as a target with a complementary site in a favorable context for stable pairing with miR-278-3p. Death-associated inhibitor of apoptosis 1 (Diap1) is predicted as a target of miR-278-3p. The downregulation of miR-278 in larvae fed on *G. physocarpus* might be involved in the upregulation of *Kibra*, *expanded* and *Diap1* and hence in deceleration of larval growth when fed on this host.

The other miRNA that was identified in both contrasts involving GP is miR-novel37-5p. This miRNA increased its expression in larvae fed on *G. physocarpus*. Among its nine targets which decreased their expression, CDK5 regulatory subunit–associated protein 1 (CDK5RAP1) is the one more clearly related to nutrient deprivation. CDK5RAP1 inhibits the activity of cyclin-dependent kinase 5 (CDK5) (Ching, Pang, Lam, Qi, & Wang, 2002). Furthermore, inhibition of CDK5 prevents the downregulation of insulin gene expression that is normally found in hyperglycemic beta cells (Ubeda, Rukstalis, & Habener, 2006). Hence, miR-novel37 supports the notion of nutrient deprivation in larvae fed on *G. physocarpus* by indirectly regulating the gene expression of insulin.

Nevertheless, none of the miRNAs that differed in expression between the *Asclepias* had protein-coding targets in the opposite direction in this same contrast. This is not surprising because there were only 66 genes could be classified as coherent targets of these miRNAs. Altogether, protein-coding and miRNA gene expression profiles agree with the observed lower growth rate of larvae fed on *G. physocarpus* than on *Asclepias*.

Besides nutritional restriction, growth impairment could also result from an allelochemical of *G. physocarpus* to which these larvae are more susceptible. If this was the case, then I would expect to see overexpression of detoxification genes. Several studies have identified cytochrome P450 genes changing expression when insects feed on different hosts (de la Paz Celorio-Mancera et

al., 2012; Koenig et al., 2015; Ragland et al., 2015; Rivera-Vega, Galbraith, Grozinger, & Felton, 2017; Wybouw et al., 2015; Q. Y. Yu et al., 2016; Zhong, Li, Chen, Zhang, & Li, 2017). This group of genes was not enriched in the differentially expressed genes. Cytochromes P450 are only the first step of three phases detoxification process: I) metabolization, II) conjugation and III) translocation (Brattsten, 1988). Neither phase II genes were enriched in any contrast, but ATP-binding cassette (ABC) transporters were enriched among overexpressed genes in larvae fed on *Asclepias* relative to larvae fed on *G. physocarpus*. ABC transporters were recently found to be relevant for insect herbivores response to different hosts (Bretschneider et al., 2016; Tian et al., 2017). The lack of enrichment of genes from phases I and II raises the question of what process might require only the last step of detoxification. Here it is worthwhile to highlight that phase III can imply excretion or sequestration. An interesting hypothesis is that these genes are relevant for cardenolides sequestration by the monarchs. Monarch larvae are able modulate the amount of cardenolides they sequester from their host and this amount varies between host species (Holzinger et al., 1992; Petschenka & Agrawal, 2015; Vaughan & Jungreis, 1977). Furthermore, Mdr49, one of the differentially expressed ABC transporters, has been recently characterized as relevant for alleviating the toxic effects of one cardenolide in *D. melanogaster* (Groen et al. 2017). Hence, monarch larvae could modulate cardenolide absorption through ABC transporters. Although some of the ABC transporters differentially expressed could be related to cardenolide sequestration, to test this hypothesis, additional experiments are needed. Such an experiment could first measure whether the host shift alters the amount of sequestered cardenolides and if this is accompanied by different abundances of the candidate ABC transporters.

Agreement between larval size and gene expression

Altogether, it seems that larvae fed on GP suffered nutritional restriction which resulted in cell growth arrest. MSc Mariana Ramírez measured larval size of larvae similarly fed for 24 hours in these hosts and this data reflects that larvae grew equally well on both *Asclepias*, but barely grew when fed on *G. physocarpus*

(González-De-la-Rosa, Ramírez Ramirez Loustalot-Laclette, *et al*. *manuscript in preparation*). Larvae consuming fewer leaves or leaves having fewer nutrients or bigger amounts of allelochemicals or more potent ones or a combination of these, all could cause retarded larval growth. Future experiments could test if larvae consume less *G. physocarpus* leaves after feeding on *A. curassavica* by measuring the mass consumed in each diet. The biochemical venue (i.e. nutrients and allelochemicals) could be probed by measuring growth of larvae fed on *A. curassavica* leaves smeared with extracts of *G. physocarpus* leaves (Thorsteinson, 1953). Either result would throw light on what causes that larvae fed on *G. physocarpus* after feeding of *A. curassavica* barely grow while under different experimental conditions *D. plexippus* larvae fed on *G. physocarpus* have a high survival rates to adulthood on this host (Tao et al., 2016). An important difference between experimental settings may be that the larvae we studied were initially reared on *A. curassavica* while those of Tao *et al.* (2016) were reared on a single species from egg hatching to pupation. The difference might alter host preference or, more appropriately, rejection. Jermy, Hanson, & Dethier (1968) observed that, in *Manduca sexta* and *Heliothis zea*, host preference can be induced by exposure to the host after the larva had previously fed on artificial diet. *Helicoverpa armigera* larvae consumed greater amounts of hot pepper than cotton leaves if they were previously fed on hot pepper while they consumed more cotton leaves if previously reared on cotton (Hu et al., 2018). I thus favor the hypothesis that larvae originally fed on *A. curassavica* consumed few *G. physocarpus* leaves because they find them less appealing than they would if they fed solely on *G. physocarpus*. The detrimental effect of allelochemicals or fewer available nutrients cannot be ruled out. I consider that the gene expression differences I observe at least partially reflect the effect of starvation. Nonetheless, larvae fed on *G. physocarpus* did grow at least a little implying that their gene expression profiles may capture the different digestion and detoxification mechanisms used to cope with the host-specific defenses. Sequencing whole larva cannot detect tissue specific gene expression differences. In *M. sexta* larvae, more genes were differentially expressed in response to host shift when comparing only the gut's transcriptomes than when assessing whole body gene expression (Koenig et al., 2015). Hence, we might have been unable to detect tissue specific gene expression differences important

to host use. Perhaps the few genes differentially expressed between the distinct *Asclepias* diets reflect that only a handful of genes are involved in consuming these different hosts. However, given the high within group variability, which I am unable to attribute to any particular factor, it is possible that several genes involved in the host-dependent response remain unidentified. Many targets of the diet-dependent differentially expressed miRNAs likely remain unidentified due to high within diet variability. The few protein-coding genes identified agree with previous studies of herbivorous insect host shifts, implicating metabolism and detoxification genes as relevant for this process (Koenig et al., 2015; Rivera-Vega et al., 2017; Silva-Brandão et al., 2017; Wybouw et al., 2015).

CONCLUSION

My study is the first to identify miRNAs and mRNAs likely involved in the interaction between monarch caterpillars and milkweeds. To tackle this question, I annotated a variety of non-coding genes, with a focus on miRNAs, and protein-coding gene untranslated regions in the monarch genome. As a result, I enriched the genomic resources available for studying the monarch butterfly. As expected, phylogenetically close hosts induced fewer expression differences compared to the most distant host, *G. physocarpus*. The analysis of protein-coding genes highlights, in larvae fed on *G. physocarpus*, the downregulation of genes involved in cell cycle progression and growth. Critical genes of the insulin signaling pathway suggest that this lack of growth might be due to a lack of nutrients. In agreement with this, miR-278-3p, previously reported to link energy homeostasis with growth, was robustly upregulated in this condition. In this thesis, I point out a complementary set of targets, corresponding to the hippo signaling pathway, that may link miR-278 with growth regulation. Several genes are likely involved in coping with the different plant nutrients and defenses rather than starvation. The ABC transporters I identified as likely involved in host response could be further tested for their role in cardenolides sequestration. Altogether, this study identified detoxification and metabolism genes possibly involved in the monarch-milkweed interaction. Characterizing the genetic diversity of these loci in monarchs could provide insights in the coevolutionary relationship between milkweeds and monarchs and, consequently, orient conservation efforts in the context of host management.

REFERENCES

Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, *4*, e05005. http://doi.org/10.7554/eLife.05005

Agarwal, V., Subtelny, A. O., Thiru, P., Ulitsky, I., & Bartel, D. P. (2018). Predicting microRNA targeting efficacy in *Drosophila*. *Genome Biology*, *19*(1), 152. http://doi.org/10.1101/198689

Agrawal, A. A., Ali, J. G., Rasmann, S., & Fishbein, M. (2015). Macroevolutionary trends in the defense of milkweeds against monarchs. In K. S. Oberhauser, K. R. Nail, & S. Altizer (Eds.), *Monarchs in a changing world: Biology and Conservation of an Iconic Butterfly* (p. 47). Ithaca, NY: Cornell University Press.

Agrawal, A. A., & Konno, K. (2009). Latex: A Model for Understanding Mechanisms, Ecology, and Evolution of Plant Defense Against Herbivory. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 311–331. http://doi.org/10.1146/annurev.ecolsys.110308.120307

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. http://doi.org/10.1111/j.1469-8137.2011.04049.x

Alcon-Giner, C., Caim, S., Mitra, S., Ketskemety, J., Wegmann, U., Wain, J., … Hall, L. J. (2017). Optimisation of 16S rRNA gut microbiota profiling of extremely low birth weight infants. *BMC Genomics*, *18*(1), 1–15. http://doi.org/10.1186/s12864-017-4229-x

Alexa, A., & Rahnenfuhrer, J. (2016). topGO: Enrichment analysis for Gene Ontology. R package version 2.28.0. *Bioconductor*.

Allikmets, R., Gerrard, B., Hutchinson, a, & Dean, M. (1996). Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database. *Human Molecular Genetics*, *5*(10), 1649–1655. http://doi.org/DOI 10.1093/hmg/5.10.1649

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

Anfossi, S., Babayan, A., Pantel, K., & Calin, G. A. (2018). Clinical utility of circulating non-coding RNAs — an update. *Nature Reviews Clinical Oncology*. http://doi.org/10.1038/s41571-018-0035-x

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. http://doi.org/10.1038/75556

Axtell, M. J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA (New York, N.Y.)*, *19*(6), 740–51. http://doi.org/10.1261/rna.035279.112

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell*, *173*(1), 20–51. http://doi.org/10.1016/j.cell.2018.03.006

Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., … Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. http://doi.org/10.1093/nar/gkw1099

Bhayani, M. K., Calin, G. A., & Lai, S. Y. (2012). Functional relevance of miRNA* sequences in human disease. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *731*(1–2), 14–19. http://doi.org/10.1016/j.mrfmmm.2011.10.014

Biggs, W. H., Meisenhelder, J., Hunter, T., Cavenee, W. K., & Arden, K. C. (1999). Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(13), 7421–6. http://doi.org/10.1073/pnas.96.13.7421

Blackwell, W. H. (1964). Synopsis of the 23 Species of *Asclepias* (Asclepiadaceae) in Tamaulipas and Nuevo Leon Including Two New Species, *Asclepias Bifida* and *Asclepias Prostrata*. *The Southwestern Naturalist*, *9*(3), 171–180. Retrieved from papers2://publication/uuid/9EEEDF12-E9DC-407A-93CF-6E42747FCEC7

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. http://doi.org/10.1093/bioinformatics/btu170

Brattsten, L. B. (1988). Enzymic adaptations in leaf-feeding insects to host-plant allelochemicals. *Journal of Chemical Ecology*, *14*(10), 1919–1939. http://doi.org/10.1007/BF01013486

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B., & Cohen, S. M. (2003). *bantam* encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in Drosophila. *Cell*, *113*(1), 25–36. http://doi.org/10.1016/S0092-8674(03)00231-9

Bretschneider, A., Heckel, D. G., & Vogel, H. (2016). Know your ABCs: Characterization and gene expression dynamics of ABC transporters in the polyphagous herbivore *Helicoverpa armigera*. *Insect Biochemistry and Molecular Biology*, *72*, 1–9. http://doi.org/10.1016/j.ibmb.2016.03.001

Brower, J. V. Z. (1958). Experimental studies of mimicry in some North American butterflies: Part I. The monarch, *Danaus plexippus*, and viceroy, *Limenitis archippus archippus*. *Evolution*, *12*(1), 32–47. http://doi.org/10.2307/2406091

Brower, L. P. (1969). Ecological Chemistry. *Scientific American*, *220*(2), 22–29.

Brues, C. T. (1920). The selection of food-plants by insects, with special reference to Lepidopterous larvae. *The American Naturalist*, *54*(633), 313–332. http://doi.org/10.1086/279763

Brunet, A., Bonni, A., Zigmond, M. J., Lin, M. Z., Juo, P., Hu, L. S., …

Greenberg, M. E. (1999). Akt promotes cell survival by phosphorylating and inhibiting a forkhead transcription factor. *Cell*, *96*(6), 857–868. http://doi.org/10.1016/S0092-8674(00)80595-4

Cech, T. R., & Steitz, J. A. (2014). The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell*, *157*(1), 77–94. http://doi.org/10.1016/j.cell.2014.03.008

Celorio-Mancera, M. de la P., Ahn, S.-J., Vogel, H., & Heckel, D. G. (2011). Transcriptional responses underlying the hormetic and detrimental effects of the plant secondary metabolite gossypol on the generalist herbivore *Helicoverpa armigera*. *BMC Genomics*, *12*(1), 575. http://doi.org/10.1186/1471-2164-12-575

Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D., & Blaxter, M. (2016). Lepbase: The Lepidopteran genome database. *BioRxiv*, 056994. http://doi.org/10.1101/056994

Chamberlain, S. A., & Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research*, (0), 1–29. http://doi.org/10.12688/f1000research.2-191.v2

Chang, H., Ai, D., Zhang, J., Dong, S., Liu, Y., & Wang, G. (2017). Candidate odorant binding proteins and chemosensory proteins in the larval chemosensory tissues of two closely related noctuidae moths, *Helicoverpa armigera* and *H. assulta*. *PLoS ONE*, *12*(6), 1–19. http://doi.org/10.1371/journal.pone.0179243

Chatterjee, D., Katewa, S. D., Qi, Y., Jackson, S. A., Kapahi, P., & Jasper, H. (2014). Control of metabolic adaptation to fasting by dILP6-induced insulin signaling in *Drosophila* oenocytes. *Proceedings of the National Academy of Sciences*, *111*(50), 17959–17964. http://doi.org/10.1073/pnas.1409241111

Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, *5*, 1438. http://doi.org/10.12688/f1000research.8987.2

Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. C., Spies, N., Baek, D., … Bartel, D. P. (2010). Mammalian microRNAs: Experimental evaluation of novel and previously annotated genes. *Genes and Development*, *24*(10), 992–1009. http://doi.org/10.1101/gad.1884710

Ching, Y. P., Pang, A. S. H., Lam, W. H., Qi, R. Z., & Wang, J. H. (2002). Identification of a neuronal Cdk5 activator-binding protein as Cdk5 inhibitor. *Journal of Biological Chemistry*, *277*(18), 15237–15240. http://doi.org/10.1074/jbc.C200032200

Cristino, A. S., Tanaka, E. D., Rubio, M., Piulachs, M. D., & Belles, X. (2011). Deep sequencing of organ- and stage-specific microrna s in the evolutionarily basal insect *Blattella germanica* (L.) (Dictyoptera, Blattellidae). *PLoS ONE*, *6*(4). http://doi.org/10.1371/journal.pone.0019350

Dassa, E. (2011). Natural history of ABC systems: not only transporters. *Essays In Biochemistry*, *50*, 19–42. http://doi.org/10.1042/bse0500019

Davis, M. P. A., van Dongen, S., Abreu-Goodger, C., Bartonicek, N., & Enright, A. J. (2013). Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods*, *63*(1), 41–49. http://doi.org/10.1016/j.ymeth.2013.06.027

de la Paz Celorio-Mancera, M., Heckel, D. G., & Vogel, H. (2012). Transcriptional analysis of physiological pathways in a generalist herbivore: responses to different host plants and plant structures by the cotton bollworm, *Helicoverpa armigera*. *Entomologia Experimentalis et Applicata*, *144*(1), 123–133. http://doi.org/10.1111/j.1570-7458.2012.01249.x

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. http://doi.org/10.1111/mec.12440

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. http://doi.org/10.1111/j.1365-2656.2007.01305.x

Dermauw, W., & Van Leeuwen, T. (2014). The ABC gene family in arthropods: Comparative genomics and role ininsecticide transport and resistance. *Insect Biochemistry and Molecular Biology*, *45*(1), 89–110. http://doi.org/10.1016/j.ibmb.2013.11.001

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

Dethier, V. (1954). Evolution of feeding preferences in phytophagous insects. *Evolution*, *8*(1), 33–54.

Dingle, H., Zalucki, M., Rochester, W., & Armijo-Prewitt, T. (2005). Distribution of the monarch butterfly, *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae), in western North America. *Biological Journal of the Linnean Society*, *85*(4), 491–500. http://doi.org/10.1111/j.1095-8312.2005.00512.x

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. http://doi.org/10.1073/pnas.1202111109

Dussourd, D. E. (1999). Behavioral Sabotage of Plant Defense: Do Vein Cuts and Trenches Reduce Insect Exposure to Exudate? *Journal of Insect Behavior*, *12*(4), 501–515. http://doi.org/0892-7553/99/0700-0501$16.00/0

Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and Plants: A Study in Coevolution. *Evolution*, *18*(4), 586–608. http://doi.org/10.2307/2406212

Endara, M.-J., Coley, P. D., Ghabash, G., Nicholls, J. A., Dexter, K. G., Donoso, D. A., … Kursar, T. A. (2017). Coevolutionary arms race versus host defense chase in a tropical herbivore–plant system. *Proceedings of the National Academy of Sciences*, *114*(36), E7499–E7505. http://doi.org/10.1073/pnas.1707727114

Erickson, J. M. (1973). The utilization of various Asclepias species by larvae of the monarch butterfly, *Danaus plexippus*. *Psyche: A Journal of Entomology*, *80*(3), 230–244.

Etebari, K., Hussain, M., & Asgari, S. (2013). Identification of microRNAs from *Plutella xylostella* larvae associated with parasitization by *Diadegma semiclausum*. *Insect Biochemistry and Molecular Biology*, *43*(4), 309–318. http://doi.org/10.1016/j.ibmb.2013.01.004

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. http://doi.org/10.1093/bioinformatics/btw354

Ewing, B., Hillier, L., Wendl, M. C., Green, P., Ewing, B., Hillier, L., … Green, P. (1998). Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. *Genome Research*, *8*(3), 175–185. http://doi.org/10.1101/gr.8.3.175

Felton, G. W., Chung, S. H., Hernandez, M. G. E., Louis, J., Peiffer, M., & Tian, D. (2014). Herbivore oral secretions are the first line of protection against plant-induced defences. In *Annual Plant Reviews: Insect-Plant Interactions* (Vol. 47, pp. 37–76). Wiley Online Library. http://doi.org/10.1002/9781118472507.ch2

Fersht, A. R., & Knill-Jones, J. W. (1981). DNA polymerase accuracy and spontaneous mutation rates: frequencies of purine.purine, purine.pyrimidine, and pyrimidine.pyrimidine mismatches during DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, *78*(7), 4251–5. http://doi.org/10.1073/pnas.78.7.4251

Fishbein, M., Chuba, D., Ellison, C., Mason-Gamer, R. J., & Lynch, S. P. (2011). Phylogenetic Relationships of *Asclepias* (Apocynaceae) Inferred from Non-Coding Chloroplast DNA Sequences. *Systematic Botany*, *36*(4), 1008–1023. http://doi.org/10.1600/036364411x605010

Fishbein, M., Juárez-jaimes, V., & Alvarado-, L. O. (2008). Resurrection of *Asclepias schaffneri* (Apocynaceae, Asclepiadoideae), a Rare, Mexican Milkweed, *55*(1), 69–75.

Flockhart, D. T. T., Wassenaar, L. I., Martin, T. G., Hobson, K. a, Wunder, M. B., & Norris, D. R. (2013). Tracking multi-generational colonization of the breeding grounds by monarch butterflies in eastern North America. *Proceedings of the Royal Society B: Biological Sciences Royal Society*, *280*(1768), 20131087. http://doi.org/10.1098/rspb.2013.1087

Forister, M. L., Novotny, V., Panorska, A. K., Baje, L., Basset, Y., Butterill, P. T., … Dyer, L. A. (2015). The global distribution of diet breadth in insect herbivores. *Proceedings of the National Academy of Sciences*, *112*(2), 442–447. http://doi.org/10.1073/pnas.1423042112

Frick, C., & Wink, M. (1995). Uptake and sequestration of ouabain and other cardiac glycosides in Danaus plexippus (Lepidoptera: Danaidae): Evidence for a carrier-mediated process. *Journal of Chemical Ecology*,

21(5), 557–575. http://doi.org/10.1007/BF02033701

Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, *40*(1), 37–52. http://doi.org/10.1093/nar/gkr688

Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, *19*(1), 92–105. http://doi.org/10.1101/gr.082701.108

Futuyma, D. J., & Agrawal, A. A. (2009). Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences*, *106*(43), 18054–18061. http://doi.org/10.1073/pnas.0904106106

Gatehouse, J. A. (2002). Plant resistance towards insect herbivores: a dynamic interaction. *New Phytologist*, *156*(2), 145–169. http://doi.org/10.1046/j.1469-8137.2002.00519.x

Granata, Ilaria, Sangiovanni, M., Thind, A. S., & Guarracino, M. R. (2017). From trash to treasure: detecting unexpected contamination in unmapped NGS data. *PeerJ PrePrints*, 10–12. http://doi.org/10.7287/peerj.preprints.3230v1

Green, T. R., & Ryan, C. A. (1972). Wound-induced proteinase-inhibitors in plant leaves: A possible defence mechanism against insects. *Science*, *175*(1967), 776–777. http://doi.org/10.1126/science.175.4023.776

Gremme, G., Steinbiss, S., & Kurtz, S. (2013). Genome tools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *10*(3), 645–656. http://doi.org/10.1109/TCBB.2013.68

Griffiths-Jones, S., Saini, H. K., Van Dongen, S., & Enright, A. J. (2008). miRBase: Tools for microRNA genomics. *Nucleic Acids Research*, *36*(SUPPL. 1), 154–158. http://doi.org/10.1093/nar/gkm952

Grimson, A., Farh, K. K. H., Johnston, W. K., Garrett-Engele, P., Lim, L. P., & Bartel, D. P. (2007). MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. *Molecular Cell*, *27*(1), 91–105. http://doi.org/10.1016/j.molcel.2007.06.017

Groen, S. C., LaPlante, E. R., Alexandre, N. M., Agrawal, A. A., Dobler, S., & Whiteman, N. K. (2017). Multidrug transporters and organic anion transporting polypeptides protect insects against the toxic effects of cardenolides. *Insect Biochemistry and Molecular Biology*, *81*, 51–61. http://doi.org/10.1016/j.ibmb.2016.12.008

Guerra-Assunção, J. A., & Enright, A. J. (2010). MapMi: automated mapping of microRNA loci. *BMC Bioinformatics*, *11*, 133. http://doi.org/10.1186/1471-2105-11-133

Haas, B. J., Delcher, A. L., Mount S.M., S. M., Wortman, J. R., Smith, R. K., Hannick, L. I., … White, O. (2003). Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids*

Research, *31*(19), 5654–5666. http://doi.org/10.1093/nar/gkg770

Han, J., Lee, Y., Yeom, K. H., Nam, J. W., Heo, I., Rhee, J. K., … Kim, V. N. (2006). Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex. *Cell*, *125*(5), 887–901. http://doi.org/10.1016/j.cell.2006.03.043

Harvey, K. F., Pfleger, C. M., & Hariharan, I. K. (2003). The *Drosophila* Mst ortholog, *hippo*, restricts growth and cell proliferation and promotes apoptosis. *Cell*, *114*(4), 457–467. http://doi.org/10.1016/S0092-8674(03)00557-9

Henry, L. M., Peccoud, J., Simon, J. C., Hadfield, J. D., Maiden, M. J. C., Ferrari, J., & Godfray, H. C. J. (2013). Horizontally transmitted symbionts and host colonization of ecological niches. *Current Biology*, *23*(17), 1713–1717. http://doi.org/10.1016/j.cub.2013.07.029

Hipfner, D. R., Weigmann, K., & Cohen, S. M. (2002). The *bantam* gene regulates *Drosophila* growth. *Genetics*, *161*(4), 1527–1537. http://doi.org/citeulike-article-id:10029271

Holzinger, F., Frick, C., & Wink, M. (1992). Molecular basis for the insensitivity of the Monarch (*Danaus plexippus*) to cardiac glycosides. *FEBS Letters*, *314*(3), 477–480. http://doi.org/10.1016/0014-5793(92)81530-Y

Howe, G. A., & Jander, G. (2008). Plant Immunity to Insect Herbivores. *Annual Review of Plant Biology*, *59*(1), 41–66. http://doi.org/10.1146/annurev.arplant.59.032607.092825

Hu, P., Li, H. L., Zhang, H. F., Luo, Q. W., Guo, X. R., Wang, G. P., … Yuan, G. (2018). Experience-based mediation of feeding and oviposition behaviors in the cotton bollworm: *Helicoverpa armigera* (Lepidoptera: Noctuidae). *PLoS ONE*, *13*(1), 5–8. http://doi.org/10.1371/journal.pone.0190401

Jermy, T., Hanson, F. E., & Dethier, V. G. (1968). Induction of specific food preference in lepidopterous larvae. *Entomologia Experimentalis et Applicata*, *11*(2), 211–230. http://doi.org/10.1111/j.1570-7458.1968.tb02046.x

Jiang, H., Lei, R., Ding, S. W., & Zhu, S. (2014). Skewer: A fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics*, *15*(1), 1–12. http://doi.org/10.1186/1471-2105-15-182

Jorgensen, P. L., Håkansson, K. O., & Karlish, S. J. D. (2003). Structure and Mechanism of Na,K-ATPase: Functional Sites and Their Interactions. *Annual Review of Physiology*, *65*(1), 817–849. http://doi.org/10.1146/annurev.physiol.65.092101.142558

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., … Petrov, A. I. (2017). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research*, (November), 1–8. http://doi.org/10.1093/nar/gkx1038

Karamysheva, Z., Díaz-Martínez, L. A., Warrington, R., & Yu, H. (2015). Graded requirement for the spliceosome in cell cycle progression. *Cell Cycle*,

14(12), 1873–1883. http://doi.org/10.1080/15384101.2015.1039209

Kessler, A., & Baldwin, I. T. (2001). Defensive Function of Herbivore-Induced Plant Volatile Emission in Nature. *Science*, *291*(March), 2141. http://doi.org/10.1126/science.291.5511.2141

Kessler, A., & Baldwin, I. T. (2002). Plant responses to insect herbivory: the emerging molecular analysis. *Annual Review of Plant Biology*, *53*(1), 299–328. http://doi.org/10.1146/annurev.arplant.53.100301.135207

Khvorova, A., Reynolds, A., & Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell*, *115*(2), 209–216. http://doi.org/10.1016/S0092-8674(03)00801-8

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and accurate classificaton of metagenomic sequences. *Genome Research*, *26*(1721–1729), 1–9. http://doi.org/10.1101/054965

Koenig, C., Bretschneider, A., Heckel, D. G., Grosse-Wilde, E., Hansson, B. S., & Vogel, H. (2015). The plastic response of *Manduca sexta* to host and non-host plants. *Insect Biochemistry and Molecular Biology*, *63*, 72–85. http://doi.org/10.1016/j.ibmb.2015.06.001

Kops, G. J., & Burgering, B. M. (1999). Forkhead transcription factors: new insights into protein kinase B (c-akt) signaling. *Journal of Molecular Medicine (Berlin, Germany)*, *77*(9), 656–65.

Kozomara, A., & Griffiths-Jones, S. (2014). MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(D1), 68–73. http://doi.org/10.1093/nar/gkt1181

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108. http://doi.org/10.1093/nar/gkm160

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., & Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science (New York, N.Y.)*, *294*(5543), 853–8. http://doi.org/10.1126/science.1064921

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. http://doi.org/10.1186/gb-2009-10-3-r25

Lau, N. C., Lim, L. P., Weinstein, E. G., & Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, *294*(5543), 858–62. http://doi.org/10.1126/science.1065062

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*, R29. http://doi.org/10.1186/gb-2014-15-2-r29

Leaman, D., Po, Y. C., Fak, J., Yalcin, A., Pearce, M., Unnerstall, U., … Gaul, U. (2005). Antisense-mediated depletion reveals essential and specific

functions of microRNAs in *Drosophila* development. *Cell*, *121*(7), 1097–1108. http://doi.org/10.1016/j.cell.2005.04.016

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. http://doi.org/10.1016/0092-8674(93)90529-Y

Lee, Y., Jeon, K., Lee, J.-T., Kim, S., & Kim, V. N. (2002). MicroRNA maturation: stepwisee processing and subcellular localization. *The EMBO Journal*, *21*(17), 4663–4670. http://doi.org/10.1093/emboj/cdf476

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. http://doi.org/10.1111/j.1365-2656.2011.01901.x

Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., & Burge, C. B. (2003). Prediction of Mammalian MicroRNA Targets. *Cell*, *115*(7), 787–798. http://doi.org/10.1016/S0092-8674(03)01018-3

Li, J., Cai, Y., Ye, L., Wang, S., Che, J., You, Z., … Zhong, B. (2014). MicroRNA expression profiling of the fifth-instar posterior silk gland of *Bombyx mori*. *BMC Genomics*, *15*(1), 1–11. http://doi.org/10.1186/1471-2164-15-410

Liang, H., & Li, W. H. (2009). Lowly expressed human MicroRNA genes evolve rapidly. *Molecular Biology and Evolution*, *26*(6), 1195–1198. http://doi.org/10.1093/molbev/msp053

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. http://doi.org/10.1093/bioinformatics/btt656

Liu, R., Holik, A. Z., Su, S., Jansz, N., Chen, K., Leong, H. S. an, … Ritchie, M. E. (2015). Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research*, *43*(15), e97. http://doi.org/10.1093/nar/gkv412

Lowe, T. M., & Eddy, S. R. (1996). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research*, *25*(5), 955–964. http://doi.org/10.1093/nar/25.5.0955

Lozano, J., Montañez, R., & Belles, X. (2015). MiR-2 family regulates insect metamorphosis by controlling the juvenile hormone signaling pathway. *Proceedings of the National Academy of Sciences*, *112*(12), 201418522. http://doi.org/10.1073/pnas.1418522112

Lucas, K. J., Zhao, B., Liu, S., & Raikhel, A. S. (2015). Regulation of physiological processes by microRNAs in insects. *Current Opinion in Insect Science*, *11*, 1–7. http://doi.org/10.1016/j.cois.2015.06.004

Marco, A., Hooks, K. B., & Griffiths-Jones, S. (2012). Evolution and function of the extended miR-2 microRNA family. *RNA Biology*, *9*(3), 242–248. http://doi.org/10.4161/rna.19160

Marr, M. T., D'Alessio, J. A., Puig, O., & Tjian, R. (2007). IRES-mediated functional coupling of transcription and translation amplifies insulin receptor feedback. *Genes and Development*, *21*(2), 175–183. http://doi.org/10.1101/gad.1506407

Matylla-Kulinska, K., Tafer, H., Weiss, A., & Schroeder, R. (2014). Functional repeat-derived RNAs often originate from retrotransposon-propagated ncRNAs. *Wiley Interdisciplinary Reviews: RNA*, *5*(5), 591–600. http://doi.org/10.1002/wrna.1243

Meng, Z., Moroishi, T., & Guan, K. (2016). Mechanisms of Hippo pathway regulation. *Genes and Development*, *30*(1), 1–17. http://doi.org/10.1101/gad.274027.115.1/2

Morreale, F. E., & Walden, H. (2016). Types of Ubiquitin Ligases. *Cell*, *165*(1), 248–248.e1. http://doi.org/10.1016/j.cell.2016.03.003

Nairz, K., Rottig, C., Rintelen, F., Zdobnov, E., Moser, M., & Hafen, E. (2006). Overgrowth caused by misexpression of a microRNA with dispensable wild-type function. *Developmental Biology*, *291*(2), 314–324. http://doi.org/10.1016/j.ydbio.2005.11.047

Nallu, S., Hill, J. A., Don, K., Sahagun, C., Zhang, W., Meslin, C., … Kronforst, M. R. (2018). The molecular genetic basis of herbivory between butterflies and their host plants. *Nature Ecology & Evolution*. http://doi.org/10.1038/s41559-018-0629-9

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, *29*(22), 2933–2935. http://doi.org/10.1093/bioinformatics/btt509

Ninova, M., Ronshaugen, M., Griffiths-jones, S., & Griffiths-jones, S. a M. (2014). Fast-evolving microRNAs are highly expressed in the early embryo of *Drosophila virilis*. *Rna*, 360–372. http://doi.org/10.1261/rna.041657.113.

Oberhauser, K., Elmquist, D., Perilla-López, J. M., Gebhard, I., Lukens, L., & Stireman, J. (2017). Tachinid Fly (Diptera: Tachinidae) Parasitoids of *Danaus plexippus* (Lepidoptera: Nymphalidae). *Annals of the Entomological Society of America*, *110*(6), 536–543. http://doi.org/10.1093/aesa/sax048

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., & Kanehisa, M. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. http://doi.org/10.1093/nar/27.1.29

Ogawa, H., Shinoda, T., Cornelius, F., & Toyoshima, C. (2009). Crystal structure of the sodium-potassium pump (Na+,K+-ATPase) with bound potassium and ouabain. *Proceedings of the National Academy of Sciences*, *106*(33), 13742–13747. http://doi.org/10.1073/pnas.0907054106

Passador-Gurgel, G., Hsieh, W. P., Hunt, P., Deighton, N., & Gibson, G. (2007). Quantitative trait transcripts for nicotine resistance in *Drosophila melanogaster*. *Nature Genetics*, *39*(2), 264–268. http://doi.org/10.1038/ng1944

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protocols*, *11*(9), 1650–1667. http://doi.org/10.1038/nprot.2016.095

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3), 290–295. http://doi.org/10.1038/nbt.3122

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*, 1–9. http://doi.org/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. http://doi.org/10.1016/j.cois.2015.12.004

Petschenka, G., Fandrich, S., Sander, N., Wagschal, V., Boppré, M., & Dobler, S. (2013). Stepwise evolution of resistance to toxic cardenolides via genetic substitutions in the Na^+/K^+-atpase of milkweed butterflies (lepidoptera: Danaini). *Evolution*, *67*(9), 2753–2761. http://doi.org/10.1111/evo.12152

Pleasants, J. M., & Oberhauser, K. S. (2013). Milkweed loss in agricultural fields because of herbicide use: Effect on the monarch butterfly population. *Insect Conservation and Diversity*, *6*(2), 135–144. http://doi.org/10.1111/j.1752-4598.2012.00196.x

Puig, O., Marr, M. T. M., Ruhf, M. L., & Tjian, R. (2003). Control of cell number by *Drosophila* FOXO : downstream and feedback regulation of the insulin receptor pathway. *Genes & Development*, *17*(16), 2006–2020. http://doi.org/10.1101/gad.1098703.messengers

Puig, O., & Tjian, R. (2006). Nutrient availability and growth: Regulation of insulin signaling by dFOXO/FOXO1. *Cell Cycle*, *5*(5), 503–505. http://doi.org/10.4161/cc.5.5.2501

Puthiyakunnon, S., Yao, Y., Li, Y., Gu, J., Peng, H., & Chen, X. (2013). Functional characterization of three MicroRNAs of the Asian Tiger Mosquito, *Aedes albopictus*. *Parasites & Vectors*, *6*(1), 230. http://doi.org/10.1186/1756-3305-6-230

Quah, S., Breuker, C. J., & Holland, P. W. H. (2015). A Diversity of conserved and novel ovarian MicroRNAs in the Speckled Wood (*Pararge aegeria*). *PLoS ONE*, *10*(11), 1–15. http://doi.org/10.1371/journal.pone.0142243

Quah, S., Hui, J. H. L., & Holland, P. W. H. (2015). A burst of miRNA innovation in the early evolution of butterflies and moths. *Molecular Biology and Evolution*, *32*(5), 1161–1174. http://doi.org/10.1093/molbev/msv004

Ragland, G. J., Almskaar, K., Vertacnik, K. L., Gough, H. M., Feder, J. L., Hahn, D. A., & Schwarz, D. (2015). Differences in performance and transcriptome-wide gene expression associated with *Rhagoletis* (Diptera: Tephritidae) larvae feeding in alternate host fruit environments. *Molecular*

Ecology, *24*(11), 2759–2776. http://doi.org/10.1111/mec.13191

Régnier, P., & Marujo, P. E. (2013). Polyadenylation and Degradation of RNA in Prokaryotes.

Ricardo, S., & Lehmann, R. (2009). An ABC transporter controls export of a *Drosophila* germ cell attractant. *Science*, *323*(5916), 943–946. http://doi.org/10.1126/science.1166239

Rivera-Vega, L. J., Galbraith, D. A., Grozinger, C. M., & Felton, G. W. (2017). Host plant driven transcriptome plasticity in the salivary glands of the cabbage looper (*Trichoplusia ni*). *PLoS ONE*, *12*(8), 1–17. http://doi.org/10.1371/journal.pone.0182636

Ro, S., Park, C., Young, D., Sanders, K. M., & Yan, W. (2007). Tissue-dependent paired expression of miRNAs. *Nucleic Acids Research*, *35*(17), 5944–5953. http://doi.org/10.1093/nar/gkm641

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. http://doi.org/10.1093/bioinformatics/btp616

Roeske, C. N., Seiber, J. N., Brower, L. P., & Moffitt, C. M. (1976). Milkweed cardenolides and their comparative processing by monarch butterflies (*Danaus plexippus L.*). In *Biochemical Interaction Between Plants and Insects* (pp. 93–167). Springer.

Satake, S., Kawabe, Y., & Mizoguchi, A. (2000). Carbohydrate metabolism during starvation in the silkworm *Bombyx mori*. *Archives of Insect Biochemistry and Physiology*, *44*(2), 90–8. http://doi.org/10.1002/1520-6327(200006)44:2<90::AID-ARCH4>3.0.CO;2-0

Satterfield, D. A., Maerz, J. C., & Altizer, S. (2015). Loss of migratory behaviour increases infection risk for a butterfly host. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1801), 20141734–20141734. http://doi.org/10.1098/rspb.2014.1734

Satterfield, D. A., Villablanca, F. X., Maerz, J. C., & Altizer, S. (2016). Migratory monarchs wintering in California experience low infection risk compared to monarchs breeding year-round on non-native milkweed. *Integrative and Comparative Biology*, *56*(2), 343–352. http://doi.org/10.1093/icb/icw030

Schoonhoven, L. M., Van Loon, B., van Loon, J. J. A., & Dicke, M. (2005). *Insect-plant biology*. Oxford University Press on Demand.

Schwarz, D. S., Hutvágner, G., Du, T., Xu, Z., Aronin, N., & Zamore, P. D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, *115*(2), 199–208. http://doi.org/10.1016/S0092-8674(03)00759-1

Seiber, J. N., Brower, L. P., Lee, S. M., McChesney, M. M., Cheung, H. T. A., Nelson, C. J., & Watson, T. R. (1986). Cardenolide connection between overwintering monarch butterflies from Mexico and their larval food plant, *Asclepias syriaca*. *Journal of Chemical Ecology*, *12*(5), 1157–1170. http://doi.org/10.1007/BF01639002

Semmens, B. X., Semmens, D. J., Thogmartin, W. E., Wiederholt, R., López-Hoffman, L., Diffendorfer, J. E., … Taylor, O. R. (2016). Quasi-extinction risk and population targets for the Eastern, migratory population of monarch butterflies (*Danaus plexippus*). *Scientific Reports*, *6*(July), 23265. http://doi.org/10.1038/srep23265

Shah, N., Nute, M. G., Warnow, T., & Pop, M. (2018). Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics (Oxford, England)*. http://doi.org/10.1093/bioinformatics/bty833

Shen, Y., Lv, Y., Huang, L., Liu, W., Wen, M., Tang, T., … Wu, C.-I. (2011). Testing hypotheses on the rate of molecular evolution in relation to gene expression using microRNAs. *Proceedings of the National Academy of Sciences*, *108*(38), 15942–15947. http://doi.org/10.1073/pnas.1110098108

Silva-Brandão, K. L., Horikoshi, R. J., Bernardi, D., Omoto, C., Figueira, A., & Brandão, M. M. (2017). Transcript expression plasticity as a response to alternative larval host plants in the speciation process of corn and rice strains of *Spodoptera frugiperda*. *BMC Genomics*, *18*(1), 792. http://doi.org/10.1186/s12864-017-4170-z

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. http://doi.org/10.1093/bioinformatics/btv351

Simon, J. C., D'alençon, E., Guy, E., Jacquin-Joly, E., Jaquiéry, J., Nouhaud, P., … Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, *14*(6), 413–423. http://doi.org/10.1093/bfgp/elv015

Singh, S., Gupta, M., Pandher, S., Kaur, G., Rathore, P., & Palli, S. R. (2018). Selection of housekeeping genes and demonstration of RNAi in cotton leafhopper, *Amrasca biguttula biguttula* (Ishida). *PLoS ONE*, *13*(1), 1–21. http://doi.org/10.1371/journal.pone.0191116

Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. 2013--2015. *Institute for Systems Biology. Http://Repeatmasker. Org*.

Smith, G., Macias-Muñoz, A., & Briscoe, A. D. (2016). Gene Duplication and Gene Expression Changes Play a Role in the Evolution of Candidate Pollen Feeding Genes in *Heliconius Butterflies*. *Genome Biology and Evolution*, *8*(8), 2581–2596. http://doi.org/10.1093/gbe/evw180

Sodani, K., Patel, A., Kathawala, R. J., & Chen, Z.-S. (2012). Multidrug resistance associated proteins in multidrug resistance. *Chinese Journal of Cancer*, *31*(2), 58–72. http://doi.org/10.5732/cjc.011.10329

Sood, P., Krek, A., Zavolan, M., Macino, G., & Rajewsky, N. (2006). Cell-type-specific signatures of microRNAs on target mRNA expression. *Proceedings of the National Academy of Sciences*, *103*(8), 2746–2751. http://doi.org/10.1073/pnas.0511045103

Stark, A., Brennecke, J., Russell, R. B., & Cohen, S. M. (2003). Identification of

Drosophila MicroRNA targets. *PLoS Biology*, *1*(3), E60. http://doi.org/10.1371/journal.pbio.0000060

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. http://doi.org/10.7554/eLife.01096

Strauss, A. S., Wang, D., Stock, M., Gretscher, R. R., Groth, M., Boland, W., & Burse, A. (2014). Tissue-specific transcript profiling for ABC transporters in the sequestering larvae of the phytophagous leaf beetle *Chrysomela populi*. *PLoS ONE*, *9*(6). http://doi.org/10.1371/journal.pone.0098637

Sugio, A., Dubreuil, G., Giron, D., & Simon, J. C. (2015). Plant-insect interactions under bacterial influence: Ecological implications and underlying mechanisms. *Journal of Experimental Botany*, *66*(2), 467–478. http://doi.org/10.1093/jxb/eru435

Surridge, A. K., Lopez-Gomollon, S., Moxon, S., Maroja, L. S., Rathjen, T., Nadeau, N. J., … Jiggins, C. D. (2011). Characterisation and expression of microRNAs in developing wings of the neotropical butterfly *Heliconius melpomene*. *BMC Genomics*, *12*(1), 62. http://doi.org/10.1186/1471-2164-12-62

Tang, H. Y., Smith-Caldas, M. S. B., Driscoll, M. V., Salhadar, S., & Shingleton, A. W. (2011). FOXO regulates organ-specific phenotypic plasticity in *Drosophila*. *PLoS Genetics*, *7*(11). http://doi.org/10.1371/journal.pgen.1002373

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

Team, Rs. (2015). RStudio: integrated development for R. *RStudio, Inc., Boston, MA URL Http://Www. Rstudio. Com*.

Teleman, A. A., & Cohen, S. M. (2006). *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes and Development*, *20*(4), 417–422. http://doi.org/10.1101/gad.374406

Thimgan, M. S. (2006). Comparative sequence analysis and tissue localization of members of the SLC6 family of transporters in adult *Drosophila melanogaster*. *Journal of Experimental Biology*, *209*(17), 3383–3404. http://doi.org/10.1242/jeb.02328

Thogmartin, W. E., Wiederholt, R., Drum, R. G., Diffendorfer, J. E., Altizer, S., Taylor, O. R., … Lopez-hoffman, L. (2017). Monarch butterfly population decline in North America: identifying the threatening processes. *Royal Society Open Science*, *4*(9), 1–15.

Thorsteinson, A. J. (1953). The chemotactic responses that determine host specificity in an oligophagous insect (*Plutella maculipennis* (Curt.) Lepidoptera). *Canadian Journal of Zoology*, *31*(1), 52–72. http://doi.org/10.1139/z53-006

Tian, L., Song, T., He, R., Zeng, Y., Xie, W., Wu, Q., … Zhang, Y. (2017). Genome-wide analysis of ATP-binding cassette (ABC) transporters in the sweetpotato whitefly, *Bemisia tabaci*. *BMC Genomics*, *18*(1), 330. http://doi.org/10.1186/s12864-017-3706-6

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., … Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28*(5), 511–515. http://doi.org/10.1038/nbt.1621

Treangen, T. J., & Salzberg, S. L. (2013). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.*, *13*(1), 36–46. http://doi.org/10.1038/nrg3117.Repetitive

Ubeda, M., Rukstalis, J. M., & Habener, J. F. (2006). Inhibition of cyclin-dependent kinase 5 activity protects pancreatic beta cells from glucotoxicity. *Journal of Biological Chemistry*, *281*(39), 28858–28864. http://doi.org/10.1074/jbc.M604690200

Ujvari, B., Casewell, N. R., Sunagar, K., Arbuckle, K., Wüster, W., Lo, N., … Madsen, T. (2015). Widespread convergence in toxin resistance by predictable molecular evolution. *Proceedings of the National Academy of Sciences*, *112*(38), 11911–11916. http://doi.org/10.1073/pnas.1511706112

Vaughan, G. L., & Jungreis, A. M. (1977). Insensitivity of lepidopteran tissues to ouabain: Physiological mechanisms for protection from cardiac glycosides. *Journal of Insect Physiology*, *23*(5), 585–589. http://doi.org/10.1016/0022-1910(77)90052-X

Venturini, L., Caim, S., Kaithakottil, G., Mapleson, D. L., & Swarbreck, D. (2017). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *BioRxiv*, 216994. http://doi.org/10.1101/216994

Vidal, O., López-García, J., & Rendón-Salinas, E. (2014). Trends in Deforestation and Forest Degradation after a Decade of Monitoring in the Monarch Butterfly Biosphere Reserve in Mexico. *Conservation Biology*, *28*(1), 177–186. http://doi.org/10.1111/cobi.12138

Vogel, H., Musser, R. O., & Celorio-Mancera, M. (2014). Transcriptome responses in herbivorous insects towards host plant and toxin feeding. In *Annual Plant Reviews* (Vol. 46, pp. 197–233). Oxford, UK: John Wiley & Sons Ltd. http://doi.org/10.1002/9781118472507.ch6

Wei, Y., Chen, S., Yang, P., Ma, Z., & Kang, L. (2009). Characterization and comparative profiling of the small RNA transcriptomes in two phases of locust. *Genome Biology*, *10*(1), R6.

Wightman, B., Ha, I., & Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*, *75*(5), 855–862. http://doi.org/10.1016/0092-8674(93)90530-4

Wiklund, C. (1974). Oviposition preferences in Papilio machaon in relation to the host plants of the larvae. *Entomologia Experimentalis et Applicata*,

17(2), 189–198. http://doi.org/10.1111/j.1570-7458.1974.tb00335.x

Wilkens, S. (2015). Structure and mechanism of ABC transporters. *F1000Prime Reports*, *7*(February), 1–9. http://doi.org/10.12703/P7-14

Wolter, J. M., Le, H. H. T., Linse, A., Godlove, V. A., Nguyen, T. D., Kotagama, K., … Mangone, M. (2017). Evolutionary patterns of metazoan microRNAs reveal targeting principles in the let-7 and miR-10 families. *Genome Research*, *27*(1), 53–63. http://doi.org/10.1101/gr.209361.116

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3). http://doi.org/10.1186/gb-2014-15-3-r46

Woodson, R. E. (1954). The North American species of *Asclepias L. Annals of the Missouri Botanical Garden*, *41*(1), 1–211. Retrieved from http://www.biodiversitylibrary.org/page/26842766

Wu, C. T., Budding, M., Griffin, M. S., & Croop, J. M. (1991). Isolation and characterization of Drosophila multidrug resistance gene homologs. *Molecular and Cellular Biology*, *11*(8), 3940–8.

Wybouw, N., Zhurov, V., Martel, C., Bruinsma, K. A., Hendrickx, F., Grbic, V., & Van Leeuwen, T. (2015). Adaptation of a polyphagous herbivore to a novel host plant extensively shapes the transcriptome of herbivore and host. *Molecular Ecology*, *24*(18), 4647–4663. http://doi.org/10.1111/mec.13330

Yin, D., Zhou, H., Kumagai, T., Liu, G., Ong, J. M., Black, K. L., & Koeffler, H. P. (2005). Proteasome inhibitor PS-341 causes cell growth arrest and apoptosis in human glioblastoma multiforme (GBM). *Oncogene*, *24*(3), 344–354. http://doi.org/10.1038/sj.onc.1208225

Ylla, G., Fromm, B., Piulachs, M.-D., & Belles, X. (2016). The microRNA toolkit of insects. *Scientific Reports, Published Online: 24 November 2016;* | *Doi:10.1038/Srep37736*, *6*(September), 281–297. http://doi.org/10.1038/srep37736

Yu, J.-Y., Reynolds, S. H., Hatfield, S. D., Shcherbata, H. R., Fischer, K. A., Ward, E. J., … Ruohola-Baker, H. (2009). Dicer-1-dependent Dacapo suppression acts downstream of Insulin receptor in regulating cell division of *Drosophila* germline stem cells. *Development*, *136*(9), 1497–1507. http://doi.org/10.1242/dev.025999

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. http://doi.org/10.1111/mec.13826

Yuan, Z., Sun, X., Liu, H., & Xie, J. (2011). MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes. *PLoS ONE*, *6*(3). http://doi.org/10.1371/journal.pone.0017666

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. http://doi.org/10.1046/j.1365-2311.2001.00313.x

Zalucki, M. P., & Malcolm, S. B. (1999). Plant latex and first-instar monarch larval growth and survival on three North American milkweed species. *Journal of Chemical Ecology*, *25*(8), 1827–1842. http://doi.org/10.1023/A:1020929732223

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. http://doi.org/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. http://doi.org/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. http://doi.org/10.1038/nature13812

Zhang, X., Zheng, Y., Cao, X., Ren, R., Yu, X.-Q., & Jiang, H. (2015). Identification and profiling of *Manduca sexta* microRNAs and their possible roles in regulating specific transcripts in fat body, hemocytes, and midgut. *Insect Biochemistry and Molecular Biology*. http://doi.org/10.1016/j.ibmb.2014.08.006

Zhao, W., He, X., Hoadley, K. A., Parker, J. S., Hayes, D., & Perou, C. M. (2014). Comparison of RNA-Seq by poly(A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*, *15*(1), 419. http://doi.org/10.1186/1471-2164-15-419

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. http://doi.org/10.1126/science.1226630

Zhong, H., Li, F., Chen, J., Zhang, J., & Li, F. (2017). Comparative transcriptome analysis reveals host-associated differentiation in *Chilo suppressalis* (Lepidoptera: Crambidae). *Scientific Reports*, *7*(1), 13778. http://doi.org/10.1038/s41598-017-14137-x