Winter 2015

Establishment of the multicolored Asian lady beetle, Harmonia axyridis, as a model system for the evolution of phenotypic variation

Lindsay Havens
University of New Hampshire, Durham

ESTABLISHMENT OF THE MULTICOLORED ASIAN LADY BEETLE, *HARMONIA AXYRIDIS,* AS A MODEL SYSTEM FOR THE EVOLUTION OF PHENOTYPIC VARIATION

BY

LINDSAY HAVENS

BS, Millersville University, 2013

THESIS

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Master of Science

in

Genetics

December 2015

This thesis has been examined and approved in partial fulfillment of the requirements of the degree of Master of Science in Genetics by:

Thesis Director, Matthew MacManes, Ph.D. Department of Molecular Cellular & Biomedical Sciences

R. Daniel Bergeron Ph.D. Department of Computer Science

David Plachetzki, Ph.D. Department of Molecular, Cellular & Biomedical Sciences

Sandra Rehan Ph.D. Department of Biological Sciences

W. Kelley Thomas, Ph.D. Department of Molecular, Cellular, and Biomedical Sciences

On December 10, 2015

Original approval signatures are on file with the University of New Hampshire Graduate School

TABLE OF CONTENTS

Dedicated to the numerous people who have helped and supported me along this journey

ACKNOWLEDGEMENTS

First off, I would like to thank my friends and family for the complete support provided to me during this process. Obtaining this degree would have been much harder without them cheering me on.

I would also like to thank my committee and many of the faculty and staff here at UNH for helping me with this project and serving as unparalleled guides throughout this entire process.

Last, but not least, I would like to thank Matt for his immense amount of help and patience. It takes a special kind of advisor to get completely onboard with a graduate student picking ladybugs as their study species.

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

ESTABLISHMENT OF THE MULTICOLORED ASIAN LADY BEETLE, *HARMONIA AXYRIDIS*, AS A MODEL SYSTEM FOR THE EVOLUTION OF PHENOTYPIC VARIATION

by

Lindsay Havens

University of New Hampshire, December 2015

The mechanisms behind the evolution and maintenance of striking visual phenotypes are as varied as the species that display these phenotypes. Multiple study systems have been well characterized and provide critical information about the evolution of these traits. However, new study systems in which the phenotype of interest can be easily manipulated and quantified are essential to answer many questions about the functionality of core evolutionary processes. One such model is elytral spot number of the multicolored Asian lady beetle, *Harmonia axyridis* (**Chapter 1)**. I describe *Harmonia axyridis* as a potential novel model species for examining extreme phenotypic evolution. To this end, I established an in-laboratory rearing protocol for *H. axyridis* (**Chapter 2**), explored the transcriptome of both the adult and larval life stages (**Chapter 3**), and conducted an initial genome analysis (**Chapter 4**). The contents of this thesis provide a characterization of the transcriptome and a draft genome that lays a foundation for further analysis and identification of the genes responsible for the continual maintenance of spot variation in *H. axyridis.*

CHAPTER I- INTRODUCTION

The Asian multicolored ladybeetle, *Harmonia axyridis*, is an organism that displays extreme localized phenotypic variation. Known phenotypic variation comes in the form of both elytra color and spot number. Elytra color can be red, yellow, orange, or black, and spot number can vary from zero to twenty-two (Koch, 2003). While these phenotypes have been previously documented, they have not been well characterized. In addition, the potential adaptive nature and genetic mechanisms underlying these polymorphisms have yet to be elucidated. In this chapter I will:

1. Report previously documented data on the phenotypes as well as our observations on the degree of polymorphism in spot number in Harmonia.
2. Discuss potential evolutionary causes for the observed polymorphisms in the context of selection found in other organisms.
3. Describe the utility of *H. axyridis* as a model organism both to investigate the evolutionary rationale of phenotypic diversity in a novel species and to explore the genetic underpinnings of these phenotypes.

1.1 Observed Phenotypic variation in *Harmonia axyridis*

Harmonia axyridis is highly polymorphic with respect to both color and spot number. Color in *H. axyridis* varies from black to red, orange, and yellow with

spot number ranging between zero and twenty-two spots (Koch, 2003). Color morphs vary based on location; black *H. axyridis* with red spots - a state known as melanic - is an uncommon morph in North America (LaMana and Miller, 1996), but is abundant in Asia (Dobzhansky, 1933). The frequency of different colors may also be influenced temperature. A decrease in melanic *H. axyridis* in the Netherlands has been shown to be correlated with an increase in average yearly temperatures (Brakefield and de Jong, 2011).

Sexual selection may play a role in color variation in *H. axyridis.* Osawa and Nishida (1992) have remarked that female *H. axyridis* choose their mates based on melanin concentration. Their choice, however, has been shown to vary based on season and temperature. Non-melanic (red, orange, or yellow of any spot number) males have a higher frequency of mating in the spring-time, while melanic (black) males have an increased frequency of mating in the summer.

To characterize phenotypic polymorphism in local populations of *H. axyridis,* animals were collected during October 2013-2015 from sides of buildings in Durham, New Hampshire. Sex and spot number was determined and recoded for each individual (**Figures 1.1, 1.2 and 1.3**). Sex was determined using a Nikon dissection scope (insert model information), and confirmed by two independent researchers. In addition, when putatively sexed individuals were placed together, mating was frequent, which further confirmed the accuracy of sex determination using the method described in 1.3.

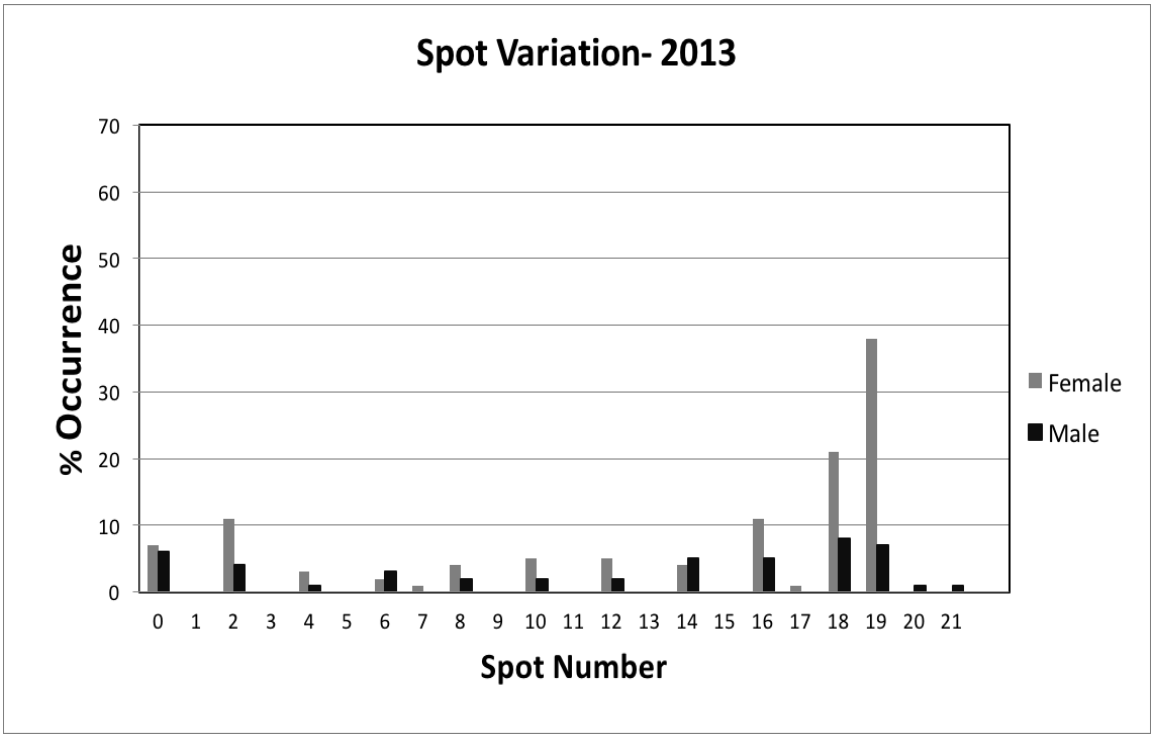

Figure 1.1: Spot variation present in 2013 in *Harmonia axyridis* collected around the University of New Hampshire in Durham, NH. 158 animals were sampled, 46 males and 112 females.

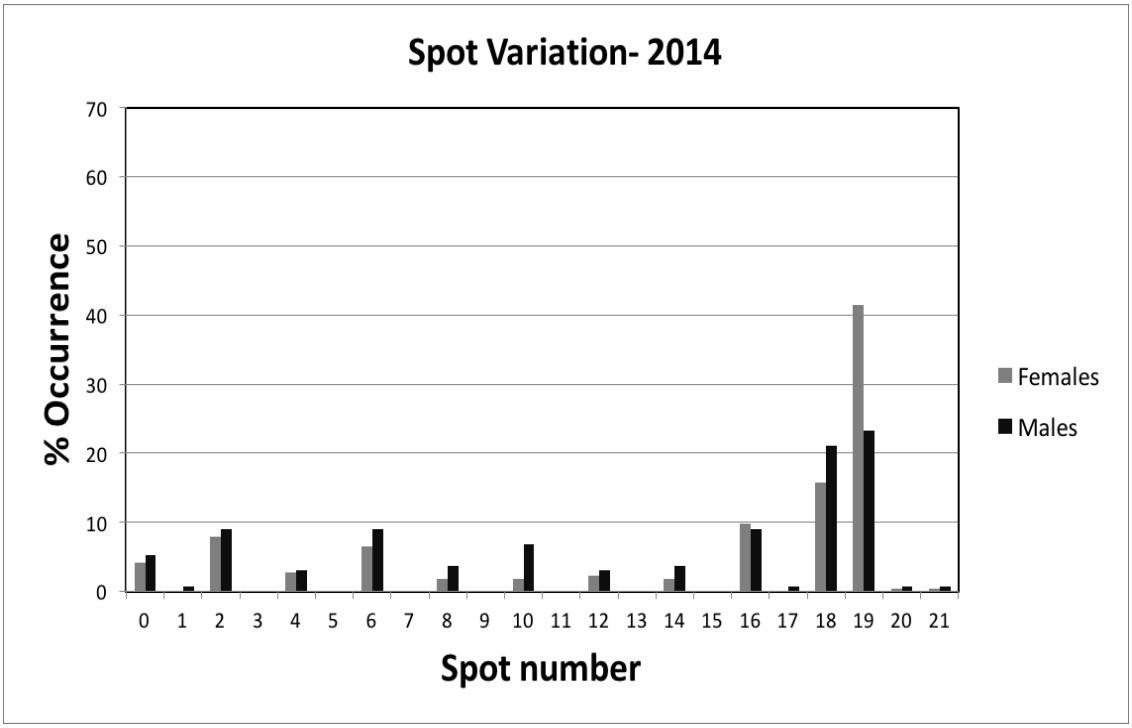

Figure 1.2: Spot variation present in 2014 in *Harmonia axyridis* collected around the University of New Hampshire in Durham, NH. 341 animals were sampled, 133 males and 208 females

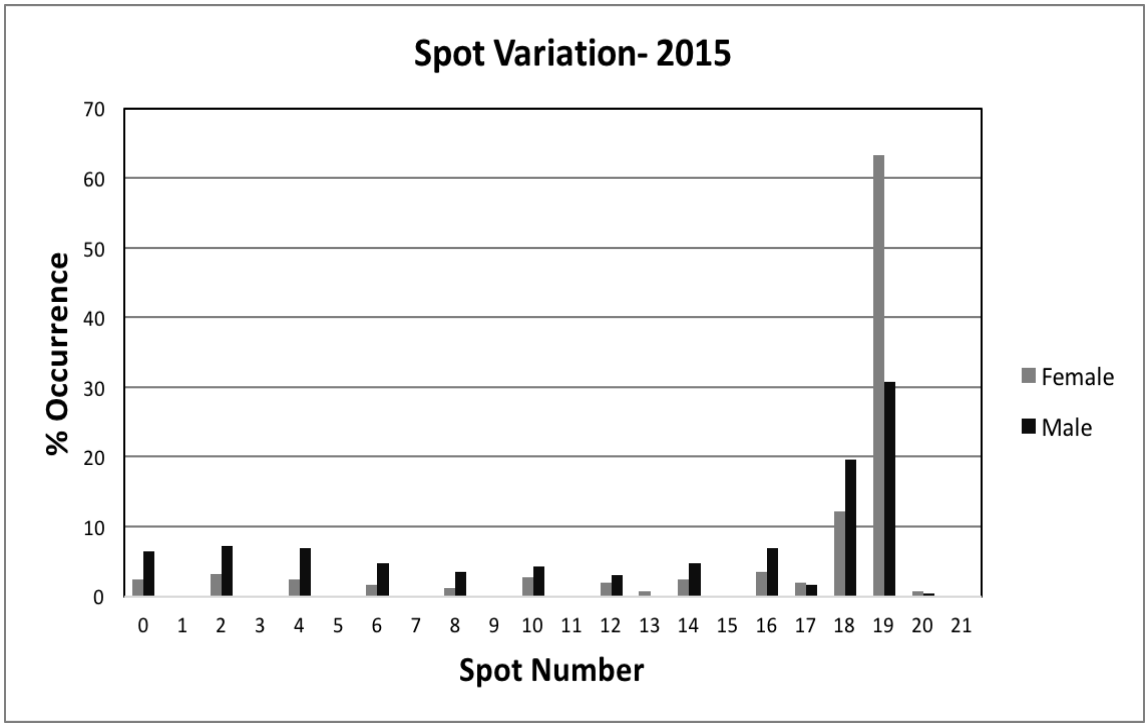

Figure 1.3: Spot variation present in 2015 in *Harmonia axyridis* collected around the University of New Hampshire in Durham, NH. 500 animals were sampled, 234 males and 256 females.

Through all observed years, the majority of specimens had 18 or 19 spots (**Figures 1.1-1.3)**. This higher spot number could prove advantageous to *H. axyridis* in colder temperatures. The increased melanin could result in better thermoregulation and survival. While a study concerning specifically spot number has not been conducted, Brakefield and de Jong (2011) found that the number of melanic individuals (black with red spots), decrease with increasing temperatures. This could indicate that elytral melanin is essential for thermal regulation processes.

1.2 Overview of evolutionary causes of phenotypic variation in nature

Phenotypic polymorphism and striking visual phenotypes have fascinated scientists for many years. The mechanisms responsible for the evolution and maintenance of these phenotypes varies as much as the species displaying them. Striking phenotypes can evolve as a result of natural or sexual selection. Sexual selection is a well characterized strategy that results in the evolution of extreme phenotypes. In contrast, three well-studied forms of natural selection that can result in striking phenotypes or phenotypic variation are aposematism, crypsis, and mimicry. I will describe these types of selection in detail below as it is currently undetermined which of these factors contribute to phenotypic

variation in *H. axyridis*. Therefore, I will evaluate the following hypotheses as evolutionary causes for phenotypic variation in *H. axyridis*:

(1) Null Hypothesis: Variation observed is a result of genetic potential for polymorphism within a population, regardless of lack of selective forces

(2) Sexual selection H1: There is selection on males by females for a particular spot number

(3) Sexual selection H2: There is selection on males by females for a particular color

(4) Natural selection H1: Evolution of phenotypic variation as a result of crypsis

(5) Natural selection H2: Phenotypic variation evolution as a result of aposematism

(6) Natural selection H3: Evolution of phenotypic variation as a result of mimicry

Phenotypic changes due to non-selection bases processes:

It is possible that the explanation for phenotypic polymorphism in *H. axyridis,* is simply due to lack of selecting pressures (either natural or sexual) to constrain these phenotypes. While this phenomenon is not well studied, a couple examples are present in the literature. Feather color variation seen in *Melospiza melodia,* song sparrows, has been linked to bacterial presence. *Bacillus licheniformis* is a feather degrading bacteria that cause feathers to

darken. This feather darkening does not confer any advantages or disadvantages to the organism (Burrt and Ichida, 2004).

Canis lupus familiaris, the common house dog, is a species that shows extensive phenotypic variation. These animals exhibit more phenotypic variation than any other terrestrial animal. While some established traits were introduced by human breeding, single-nucleotide polymorphism (SNP) analysis showed that three or fewer quantitative trait loci underpin the incredible diversity seen in *C. l. familiaris* (Boyoko, 2010).

Sexual selection

Sexual selection is a mode of natural section wherein one sex preferentially mates the other based on specific characteristics (Lande, 1981). These characteristics are under strong evolutionary control, as the presence or absence of them directly results in mating success or failure. In some extreme cases, previously unfavorable phenotypes become common throughout the population as the the fitness gained through mating can override survival value (Lande, 1981).

Sexual selection can be a result of both female or male choice. In most species, females are the choosy gender. For example, female honey bees, *Apis mellifera,* mate preferentially with larger males when presented with the option. This results in selection for a large body size. This causes the average size of males within the population to increase (Couvillon et al., 2010).

Male mate choice arises either when males invest a considerable amount of energy into offspring, or where female phenotypic variation is particularly great (Simmons, 1992, Gwynne, 1993). Male mate choice falls under two categories: cryptic, where there is a varying male response (such as varying ejaculate transfer), and male choosiness. Male choosiness differs from cryptic selection in that males openly reject mates based on a phenotypic characteristic. Male mate-choice can also be a combination of cryptic choice and choosiness (Bonduriansky, 2001).

One well defined system that utilizes both male and female mate choice is illustrated by the potbellied seahorse, *Hippocampus abdominalis*. Both males and females of this species show preferential malting, but the sexes choose mates based on different traits. Females mate based on male dissimilarity in the MH11b gene. In contrast, males choose mates based on size, preferentially mating with large females. As male *H. abdominalis* carry the young, male reproductive investment in the young is substantial (Bahr, et al., 2012).

Although, as stated in previous research described in Section 1.1, while female *Harmonia axyridis* have been shown to choose males based on elytral color, no such choice has been documented with respect to spot number. The spot variation present in male *H. axyridis* is not indicative of sexual selection (**Figures 1.1-1.3**). If female preferentially chose one phenotype over another, male spot number should have converged on the preferred number rather than displaying the large breadth of spot numbers present in the population. Because

of the high frequency of individuals with 19 spots, the argument could be made that *Harmonia* are in the beginning stages of sexual selection, although this is common to both sexes. However, this argument is not consistent with the life history strategy of these insects, as male mate choice is unlikely because males do not invest a substantial amount of energy in offspring.

Crypsis

Another mechanism by which phenotypic variation arises is through the evolution of cryptic coloration. Cryptic coloration, a survival strategy in which prey aim to avoid predator detection, can be broken down into two sub-categories – major and minor color variation (Tsurui, Honma, and Nishida, 2010). Both major and minor color variation help prey remain undetected by predators. Major variations allow prey to blend into specific environments; minor variation in phenotype can aid in prey survival by dividing predator attention among more than one phenotype (Forsman et al., 2011).

The pygmy grasshopper, *Tetrix subulata,* is an organism that utilizes both of these cryptic coloration models. *T. subulata* vary greatly in coloration; the major color morphs are black, brown and light grey. *T. subulata* found on recently burnt earth have been shown to undergo rapid evolution to more melanistic forms (Forsman et al., 2011). All of these major color variations are accompanied by minor variations in color patterns, morphology, physiology and life-history. Karpestam et al. (2014) found reduced predator effectiveness related not only

to background camouflage, but also to natural levels of minor polymorphism within the microhabitat.

Another organism that displays between-habitat phenotypic variation is the beach mouse, *Peromyscus polionotus*. The coat color varies in direct response to the environment in which they live (Vignieri et al., 2010). The lighter color morph is prevalent in sand environments, while the darker phenotype more frequently observed in field habitats. Hoesktra et al. (2006) identified a single amino acid change in melanocortin-1 receptor (Mc1r) that explains much of this phenotypic variation. Interestingly, mutations in Mc1r has also been shown to play a role in melanistic coat polymorphism in *Panthera onca* (Eizirik, 2003).

While minor cryptic color variation could explain the spot number variation present in *Harmonia axyridis*, support for a cryptic evolution hypothesis is lacking. The bright red, organge and yellow coloration of *H. axyridis* is not indicative of a predatory view evasion strategy. Animals displaying crypsis, as described above, use this strategy to blend into their natural surroundings. *H. axyridis* are predominately found in bushes and on trees (Via, 1999; Brown and Miller, 1998; Brown et al., 2011), so their bright coloration would not allow them to blend into these environments effectively.

Aposematism

Another anti-predator strategy employed across the animal kingdoms is aposematism, a tactic where a signal is used to warn predators of the distasteful or poisonous nature of prey (Santos, Coloma, and Cannatella, 2003). Bright colors of aposematic animals, such as members of the poison dart frog family, *Dendrobatidae,* serve as a conspicuous signal to predators of the inedible nature of their prey (Cadwell, 1996). Bright coloration is linked to increased predator learning, leading to a reduction of prey death (Gittleman et al., 1920). While conspicuous coloration can initially cause increased chance of death when compared to the cryptic strategy, Sword (2002) suggests that population density can offset the initial cost of coloration.

There has been much debate among scientist concerning the evolution of aposematic coloration. Fisher (1930) suggested that the evolution of distastefulness accompanying bright coloration ultimately arose through kin altruism, while Servedio (2000) used mathematical models to show that the speed of learning has a great effect on the evolution of aposematism. Servedio (2000) also found three factors essential for the evolution of aposematism without mimicry. Without quick predator learning, slow forgetting, and pervasiveness of brightly colored prey, aposematic evolution would be unlikely.

Aposematism could be a viable explanation of the evolution of striking coloration in *H. axyridis*. *H. axyridis* excrete a distasteful reflex-bleed composed

of alkaloids. While this will not kill predators, it transforms *H. axyridis* into a distasteful prey (Alam et al., 2002). The bright coloration of *H. axyridis* as well as the spot patterns present, could serve to warn predators that they are unpalatable. While the intraspecific variation in coloration and spot number would not be beneficial to serve as a warning signal, most individuals of the population display bright coloration. Therefore, the striking coloration itself is consistent with aposematism.

Mimicry

Mimicry occurs when one species evolves to resemble another. Two well-known forms of mimicry employed by organisms, through which striking warning phenotypes evolve, are Batesian and Müllerian mimicry. These two forms of mimicry vary in the palatability of the organisms involved (Fisher, 1930). Batesian mimicry, seen in the relationship between the scarlet kingsnake (*Lampropeltis elapsodies*) (the mimic) and the coral snake (*Micruroids fulvius*) (the model) is a strategy in which a palatable, or non-venomous, species mimics an unpalatable, or venomous, model (Pfennig, Harcombe, and Pfennig, 2001). The result is beneficial for the mimic, but costly for both the predator and the model. Therefore, this strategy is most successful if the mimic is relatively rare and evolves rapidly when compared to a severely unpalatable model (Fisher, 1930; Lindstrom, Alatalo and Mappes, 1997).

In the Müllerian mimicry strategy, both the mimic and the model are either unpalatable or venomous (Müller, 1879). Müllerian mimicry is advantageous for the mimic, model, and the predator predator. The mimic and the model receive added protection from predators and the predator's learning in simplified (Fisher 1930). One particularly well-characterized species that exhibits phenotypic polymorphism is found in the Neotropical butterfly genus, *Heliconius*. The color, pattern, and eyespot polymorphism seen in *Heliconius* is thought to have arose as a result of Müllerian mimicry (Flanagan et al., 2004) and the genes underlying these traits have been well characterized (Kronforst et al., 2006, Joron et al., 2006, Jones et al., 2012).

The variation present in *Harmonia axyridis* is not indicative of a mimicry strategy. Most mimics, like *M. fulvius* and within the *Heliconius* system converge on a preferred phenotype. While the 19 spot phenotype is the most common in *H. axyridis*, the variation in spots and colors does not indicate phenotypic convergence. In addition, as an invasive species, very few things prey on *H. axyridis* in its adult form. In fact, *H. axyridis* preys on all other species of ladybeetle, while the reverse predation is very rare (Hodek, van Emden, and Honêk, 2012).

Summary of phenotypic evolution in nature

To summarize, *H. axyridis* does not fit current models of sexual selection, mimicry or crypsis. Due to the distasteful nature of the reflex bleed produced by *H. axyridis* under distress, the bright coloration could serve as a warning signal for

predators, which supports aposematism as an evolutionary cause for bright coloration.

My research on *Harmonia axyridis* aims to elucidate the mechanistic links between phenotype and genotype, which will provide a unique opportunity to gain insight into the inner workings of many core evolutionary processes. While systems like the ones described above have been pioneering in exploring the evolutionary drivers of phenotypic variation, the use of novel models, especially those that can be easily manipulated, are necessary. *Harmonia axyridis* is a novel study system that processes many qualities of existing model organism while also offering several unique benefits that will be discussed below.

1.3 *Harmonia axyridis* as a potential model organism for phenotypic polymorphism

Model organisms, non-human models that are utilized to better understand biological phenomena, are fundamental to understanding the genetic foundation for observed traits. Established model organisms, such as *Drosophila melanogaster* and *Arabadopis thaliana*, share a suite of notable characteristics. Among these are: fast generation time, low cost to collect and maintain, and susceptibility of genetic modification (Leonelli and Ankeny, 2013). Many aspects of the *Harmonia axyridis* life cycle, such as (1) large egg clutch size, (2) holometabolous life stages, and (3) ease of both collection and identification of

key phenotypes, match these criteria and are conducive to its establishment as a novel model organism for the study of phenotypic evolution.

Egg clutch size

Like the *Drosophila* model system, Harmonia *axyridis* lay eggs in clutches. While the average number of eggs per clutch is 21, clutch size can be as large as 72 eggs (Hodek, van Emden, and Honêk, 2012). Ultimately, the maximum number is determined by number of ovarioles present within the ovary (Hodek, van Emden, and Honêk, 2012). This number has been seen to vary with diet. Females reared on a diet of *Acyrthosiphon pisum* laid larger clutches when compared to females reared on a diet of heterospecific and conspecific eggs (Ware et al., 2008).

Ensuring monogamous mating

To accurately follow progeny throughout generations, care needs to be taken to ensure egg clutches are a result of monogamous mating. Promiscuity is common in *H. axyridis* and has been shown to increase egg clutch size in other ladybird beetles, such as *Propylea dissecta* and *Cheilomenes sexmacuulata* (Omkar and Pervez, 2005; Bind, 2007). With respect to *H. axyridis,* Awad et al., (2015), showed that a single egg clutch could be the product of multiple mating; so desired pairs should be kept separated from the rest of the population. Strategies to ensure that the egg clutch is the product of a monogamous mating will be discussed in Chapter 2.

Hatching rate of these eggs vary with some environmental factors, such as temperature (Lombaert et al., 2008). The trophic eggs (unfertilized eggs that will not hatch) serve as a food source for siblings. In *H. axyridis*, the laying of trophic eggs is a common maternal strategy (Perry and Roitberg, 2005).

Holometabolous life stage

Harmonia axyridis are holometabolous, meaning they have complete metamorphosis between the larval, pupae, and adult stages (Hodek, van Emden, and Honêk, 2012). After emergence from the egg, all life stages are external, lending to ease of observation and collection at every phenotypic change. In the larval stage, metamorphosis between the first three instar stages (the periods between molting) are accompanied by phenotypic differences (**Figure 1.4**). Development from the first to second instar typically occurs between days 2 and 3, with little phenotypic change. Moulting between days 4 and 5 is accompanied by great phenotypic change in the third instar larvae, followed by spots appearance on the lower abdomen between days 6 and 7. The fourth instar (days 11 and 12) stops feeding and attaches to a substrate to prepare for pupation. The amount of time spent in each larval stage varies with temperature and both the quality and quantity of food available (Hodek, van Emden and Honêk, 2012).

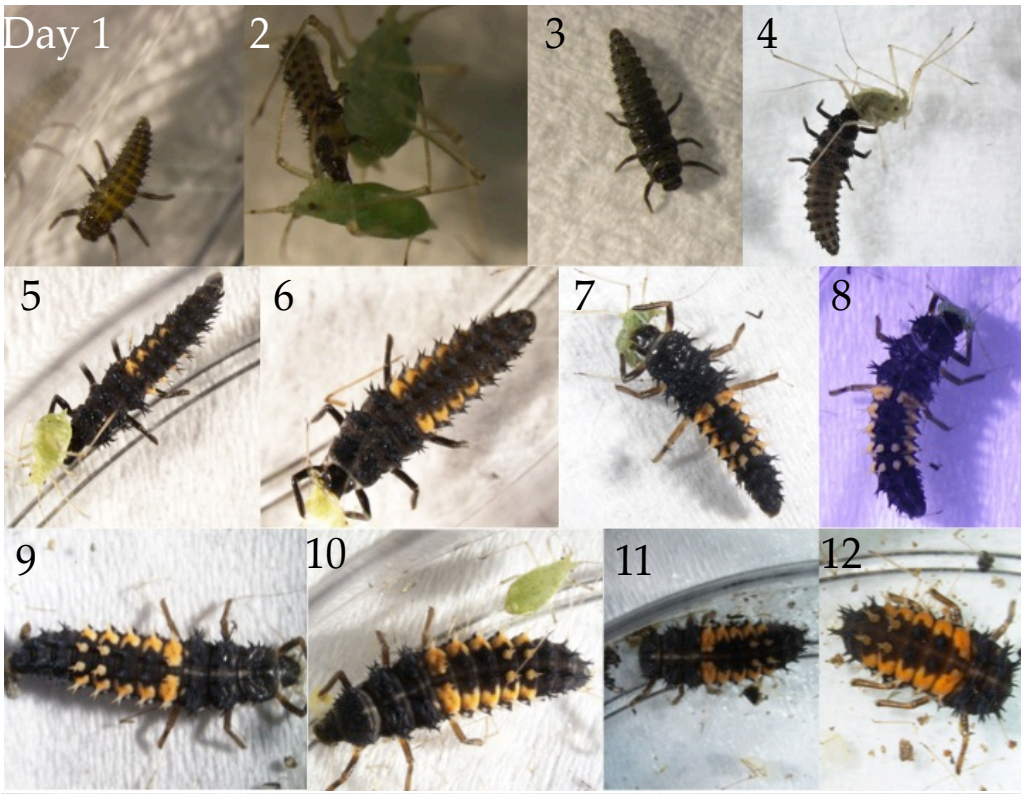

Figure 1.4: Phenotypic changes throughout the larval life stage of *Harmonia axyridis*. A single larva was followed and photographs were captured daily.

A phenotypic change also accompanies the entrance to the pupal stage. Spots appear and darken throughout the pupal cuticle development (**Figure 1.5**).

Ease of collection

Harmonia axyridis enter a state of diapause during the winter months. *H. axyridis*' diapause is facultative and the entry is regulated by environmental cues, such as day length and temperature (Hodek, van Emden, and Honêk, 2012). In many areas, *H. axyridis* seek refuge in warm areas and spend the winter months in the walls of buildings. This means during the late fall, hundreds of *H. axyridis* migrate to the sides of buildings on sun-filled days. Researchers can easily collect the insects to bring back to the laboratory setting for further analysis.

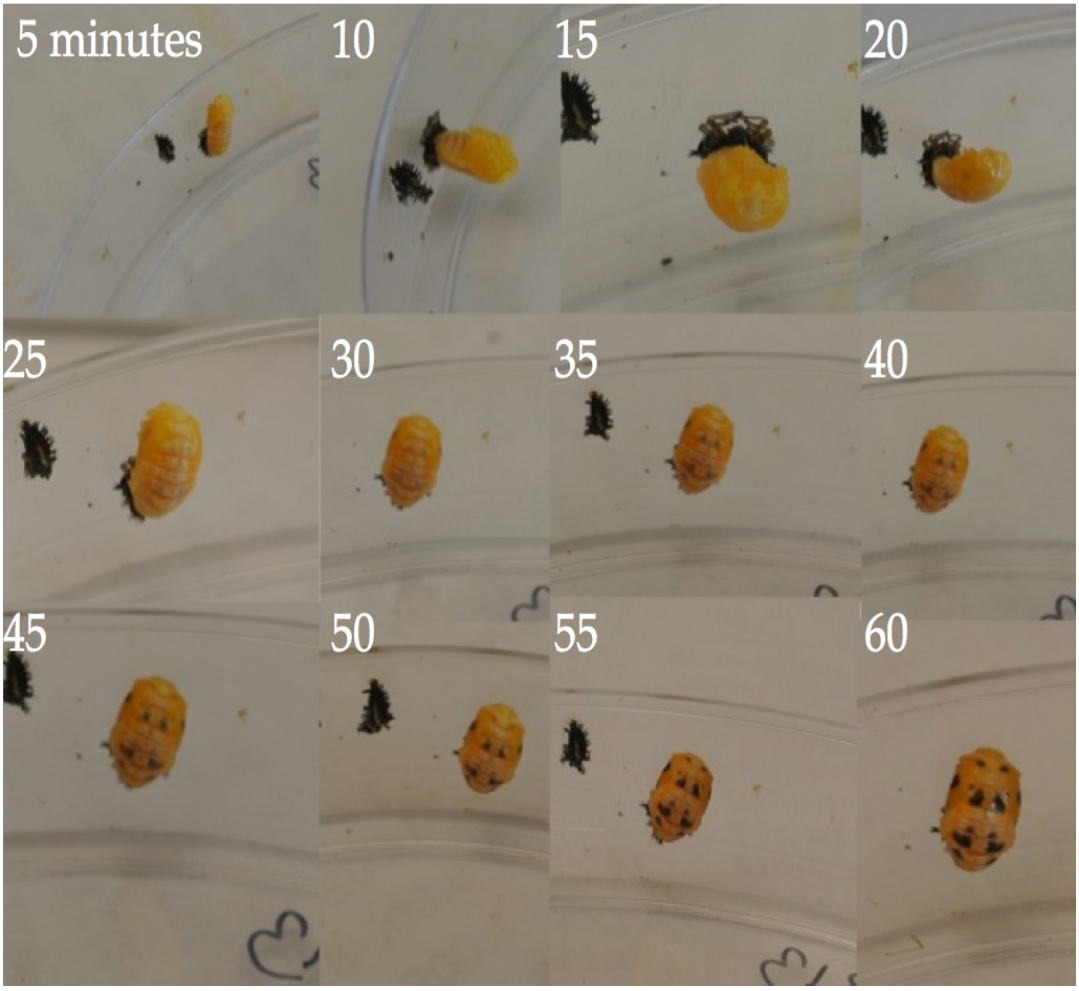

Figure 1.5: Phenotypic change accompanying pupa cuticle formation. Photographs were taken 5 minutes apart.

Easily quantified phenotype

Finally, the pigmentation phenotypes of *H. axyridis* is are both striking and easy to quantify. Spot number can be easily counted and recorded, and sex can be determined using a dissection microscope. Males have an arched second segment, while female's second segment is mostly flat (**Figure 1.6**).

Summary of Harmonia axyridis as a model species

Many characteristics of *H. axyridis* make it well suited as a novel model organism for studying phenotypic evolution. Its ease of collection and large clutch size allows for large study sample sizes. The completely external metamorphosis characteristic of changes between life stages enables study of phenotypic changes throughout development. The ease of phenotypic characterization also allows for quick and efficient classification of elytral color and spot polymorphisms.

Figure 1.6: Sex determination of *Harmonia axyridis*. Sex was determined by looking at the second segment (circled above).

1.4 Chapter Summary

Harmonia axyridis is a species that displays striking phenotypic variation in both color and spot number. Around Durham, NH (43.1339°N, 70.9264° W) the most prevalent spot numbers documented in the collection for 2013 to 2015 were years 18 and 19 spots (**Figures 1.1-1.3**). While the varying color and spot numbers present in populations of *H. axyridis* have been well characterized, the evolutionary processes that underlie the evolution of these traits is currently unknown. However, the evolutionary processes that underlie the evolution of striking phenotypes in other species is well characterized.

We evaluated multiple hypotheses for the evolutionary cause of color and spot number variation in *Harmonia axyridis.* Mimicry is an unlikely explanation as *H. axyridis* preys on local ladybeetles and has no environmental mimic. Crypsis is an unlikely hypothesis, as the bright colors displayed are not conducive to remaining hidden. Sexual selection, initialized by either males or females, is also unlikely, as the current variation seen in both sexes is not conducive to convergence on a single preferred phenotype. When considering all of the above hypotheses, aposematic evolution is the most strongly supported hypothesis for the color and spot morph evolution seen in *H. axyridis*. *H. axyridis* releases an alkaloid-containing reflex bleed that predators find distasteful. The bright colors of *H. axyridis* could serve as a signal warning predators of the unpalatable nature of their prey.

We have also described the utility of *H. axyridis* as a novel study species to investigate phenotypic polymorphisms. This species has the three traits making it an ideal model organism: large egg clutch size, ease of collection, and quantifiable color and spot numbers. In order to elucidate the genetic underpinning of phenotypic variation in *Harmonia axyridis*, establishing an in-laboratory population is necessary. To this end, the content of chapter 2 will outline how to:

1. Collect and store an initial population of *H.* axyridis
2. Establish an in-laboratory mating population and ensure monogamous mating
3. Characterize phenotypic diversity

CHAPTER II – ESTABLISMENT OF A MATING COLONY OF *HARMONIA AXYRIDIS* IN A LABORATORY SETTING

A controlled mating population is critical for establishing inheritance of phenotypic characters and elucidating the relative extent of environmental and genetic control for polyphenism. The ability to establish a mating population readily within a laboratory setting is a large benefit for using *Harmonia axyridis* as a model species. In the laboratory, controlled mating pairs can be established and environmental conditions can be manipulated.

2.1 Collection of *Harmonia axyridis*

Harmonia axyridis can be collected in type of any container and multiple insects can be stored together. Care should be taken to allow air flow into the container, but holes should not be large enough to allow insects to escape. A standard pushpin was used to puncture the lid of a 50 mL conical tube (VWR ™).

Storage

If newly collected *H. axyridis* are not immediately required, insects can be stored at 4°C for long periods of time. The cold temperature promotes the entrance into diapause and a food source is not necessary (Rake-van den Berg et al., 2012). Insects can be stored either individually in 1.5 mL Eppendorf VWR™ tubes or combined in 50 mL conical tubes. While food is not necessary for storage, it is imperative that the storage containers have holes in them to allow

air flow before placement in 4°C. While group storage is acceptable, separating out individual insects before cooling allows for easier phenotypic characterization when they are removed at room temperature.

Summer months

During the summer months, *H. axyridis* populations converge around their preferred food source, the pea aphid, *Acyrthosiphon pisum*. *A. pisum* are commonly found on alfalfa, clover and legumes, so consequently, *H. axyridis* are found in these locations (Via, 1999). In addition to these locations, *H. axyridis* can also be found on fruit trees, tobacco and wheat (Brown and Miller, 1998; Brown et al., 2011). Sweep nets can be used to capture insects from shrubs and trees, but manual collection is feasible if no sweep net is available (Elliot et al., 1991). During the summer, it is possible to collect all life stages of *H. axyridis*, although the pupal and egg stages require manual collection because they are firmly attached to leaves.

Fall months

In order to prolong life and increase survival during colder temperatures, *H. axyridis* enter a state of facultative diapause (Hodek, 2012). Winter diapause in the beetle family (Coccinellidea) is stimulated by a combination of internal and environmental cues, and is accompanied by alterations in physiology, morphology, and behavior (Denligner, 2002). In preparation for winter diapause, large swarms of *H. axyridis* aggregate on south, west, or southwest sides of light-

colored buildings (Kidd et al., 1995). As a consequence, they are easiest to capture during the early fall months, with most locations reporting aggregations occurring in October (Raak-Van den Berg et al., 2013). Insects can also be collected in the same locations as they are found during the summer months. Most specimens collected during the fall will be in the adult stage.

Winter months

Following the arrival at the aggregation site in the fall, *H. axyridis* are driven by physical factors – such as thigmotaxis, negative phototaxis, and geotaxis – to hibernate in dark, protected places (Hodek, van Emden and Honêk, 2012). As a result, many *H. axyridis* hide inside walls of houses during the winter months. *H. axyridis* is a serious indoor winter allergen because humans can become allergic to proteins excreted by them (Nakazawa et al., 2007).

Individuals will enter diapause if care is taken to capture *H. axyridis* immediately upon entrance into diapause in the winter, phenotypic ratios will be consistent with the fall collection. As specimens perish throughout the winter, this ratio is likely to change. In addition to houses, *H. axyridis* also spend their winter months in crevices of rocks as well as pine and apple trees, and they can be collected from these locations (Kidd et al., 1995).

Spring months

H. axyridis become active when ambient temperatures exceed 10 °C. In warmer climates, beetles begin to emerge from hibernation as early as mid-

February. This emergence continues over several weeks (Hodek 1961). As winter diapause ends, environmental cues stimulate mating behavior. The resulting generation emerges from the pupal stage in early April. *H. axyridis* are a multivoltine species, and the second generation emerges from the pupal stage in June. In Greece, as many as five generations of *H. axyridis* have been observed simultaneously in one location (Majerus, Strawson and Roy, 2006).

2.2 Establishing an in-laboratory mating population

Multiple factors must be considered prior to the establishment an in-laboratory mating population. Aphid populations must be established and maintained. Also, the photoperiod has to be conducive to mating. Furthermore, individual *H. axyridis* must be separated at all stages from hatching until mating to ensure monogamous matings.

Establishment of a confined aphid population

Acyrthosiphon pisum, is the preferred food source of *H. axyridis.* If care is taken to confine *A. pisum* to one location in the laboratory, a population can be established with ease, and any unwanted *A. pisum* spread can be prevented. *A. pisum* has both a winged and a wingless morph. The winged morph displays migratory behavior, while the wingless morph lives a sedentary life. The presence or absence of wings can be controlled by genetics, environment, or a combination of the two (Cailaud et al., 2002). Emergence of winged morphs can be triggered by low day light, low temperature, low food

quality and high density (Hardie and Lees, 1985). In a laboratory setting, these specific conditions can be prevalent, so the winged morph is common.

The construction of a confining habitat will prevent winged *A. pisum* spread. To construct this habitat, the following materials will be needed:

1. A full-UV range grow light
2. Four pine wood 2 X 4's
3. One 4ft by 8ft pegboard
4. One 3/4 –in X 48-in X 96-in commercial particleboard
5. One Phifer 18 X 14 Pool and Patio, 60-in by 25ft Charcoal Fiberglass Screen Wire
6. One 1 qt high-gloss white Rust-Oleum paint

In the set-up shown in **Figures 2.1** & **2.2**, the pegboard and particle boards were both cut into 4 X 4 pieces and secured with screws. The 2 X 4 pieces were used to reinforce this structure. The exposed particleboard was painted with Rust-Oleum paint to protect against moisture damage, and the grow light was set to continuously on. Finally, the peg board was covered with the fiberglass screen wire to prevent aphid escape (**Figures 2.1** & **2.2**).

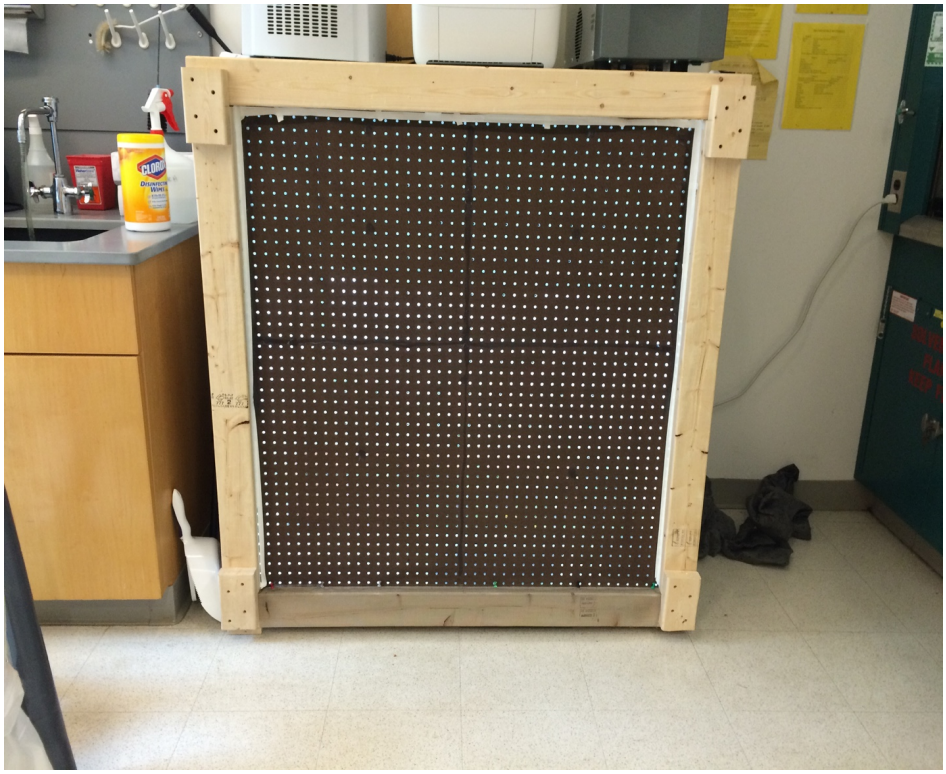

Figure 2.1: The outside of an aphid confinement habitat

Figure 2.2: The inside of an aphid containment habitat

To maintain *A. pisum* populations, fava bean, *Vicia faba*, seeds must be planted continuously. To fit into the prebuilt habitat, five Garden Treasures 6-in window boxes were filled with Miracle-Gro® Flower and Vegetable Garden Soil. Fava bean seeds were planted approximately 1 inch apart and these boxes were placed within the habitat.

Since *A. pisum* effectively destroy the plants they feed on, a frequent rotation of plants and boxes approximately twice monthly is necessary. In order to prevent insect spreading, newly growing plants should never be placed in the laboratory setting outside of the containment box.

Individual H. axyridis placement

Individuals of all life stages (mating pairs, larva, pupae, and adult) were held in 60 x 15 mm VWR ™ anaerobic dishes. Using anaerobic petri dishes is imperative as newly emerged larva are incredibly small and can easily escape aerobic petri dishes (**Figure 2.3**). Each individual *H. axyridis* should have free access to water. A slightly water-damp cotton ball can be placed in the habitat after larva reach the second instar stage. If placed in with the larva before this stage, drowning may occur. Care should be taken to reduce the amount of available space under the damp cotton ball. The tubercules on the back of all larva stages leave it susceptible to entanglement in the cotton, which will cause the larva to perish.

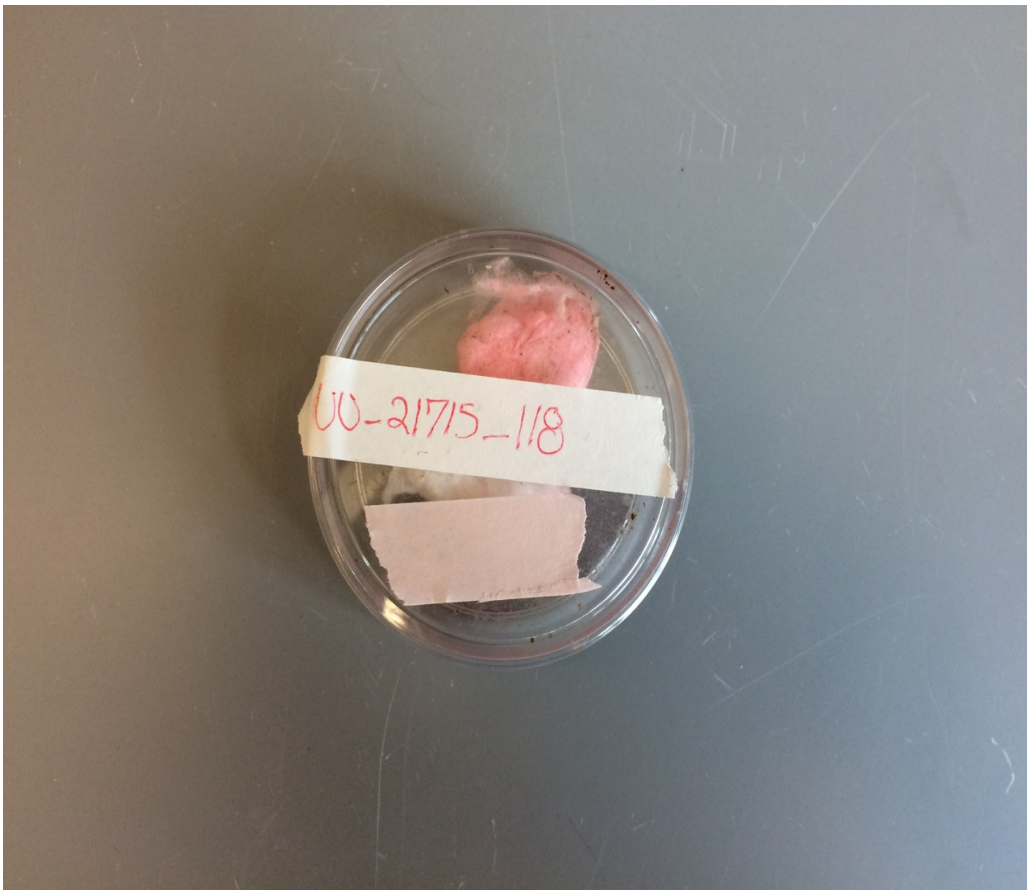

Figure 2.3: Example of an in-laboratory *H. axyridis* habitat

Cleaning of the habitat should be performed every few days. After cleaning, a new, damp cotton ball should be placed in the dish. Once the specimen reaches the pupal stage, the cotton ball can be removed to prevent molding accumulation within the habitat.

Adult *H. axyridis* eat between 15-65 aphids per day (Koch, 2003). We recommend providing a minimum of 10 aphids daily for each individual. Therefore, aphid population numbers can become the limiting factor for expanding the breeding population. If the aphid population is running low, diluted humming bird nectar and water-soaked raisins can serve as an alternative food source. Similar to the cotton ball used to provide water, a cotton ball can be dipped in humming bird nectar (such as diluted PerkyPet® nectar concentrate - www.perkypet.com), wrung until nearly dry, and placed in the *H. axyridis* environment. Raisins should be soaked in water for more than an hour. In *A. pisum* absence, mating ceases and larva development is slowed (Hodek, van Emden and Honêk, 2012), but these food alternatives will allow the *H. axyridis* to survive until the aphid population rebounds.

Photoperiod

To counteract the conditions that stimulate entrance into diapause (low temperatures and a short day length of10L:14D) a summer day length cycle should be established, and a heat lamp should be used to elevate the temperature above 10°C. If a long day length light cycle and warmer

temperatures are not established, the specimens that were collected during the fall will continue into winter diapause, and they will not mate until favorable conditions are met.

A Fluker's® turtle clamp lamp and a 75 Watt light bulb, paired with a 24-hour mechanical timer, was used to a illuminate the insect area from 6AM to 8PM (14L:10D). In addition, a ZooMed® wire cage clamp lamp was used as a heat lamp and turned on at all times. Care needs to be taken to ensure that the heat lamp is a sufficient distance from the insect area. The area needs to be heated, but the amount of condensation within the habitats needs to be minimal, as larva of all size will drown in any small amount of free liquid.

Appropriate labelling

Once eggs have been laid, the mating pair should be removed and placed into a clean, labeled dish with a large supply of aphids to prevent their consumption of the eggs. Matings will often result in more than one egg clutch. It is imperative that an appropriate labelling system is implemented to recognize siblings, as this allows the researcher to follow phenotypes through generations. The coding system used was:

Female's 1 digit color | female's 1 or 2 digit spot # | male's 1 digit color | male's 1 or 2 digit spot # __ # of day in given year eggs | 2 digit year __ Clutch letter (ABC…etc.) | larva #

Larva number was randomly assigned when removing larva from the dish containing the egg clutch. Color was assigned by eye approximation and the 1 digit colors were:

Orange = 1

Red = 3

Orange - red = 4 (1+3)

Yellow = 5

Orange - yellow = 6 (1+5)

For example, if an orange spotless female and an orange spotless male mated, and eggs were laid on March 5, 2015, the first larva from clutch A would be labelled:

1010_6415_A1

These labels were hand-written on VWR ™tape. If multiple generations are present in the insect containment area, a variety of tape colors allows efficient identification of siblings. Several tape colors are also useful if multiple experiments are being conducted concurrently. Multiple colors can also be used to quickly identify the sex of the individual when choosing mating pairs.

Larva extraction

In order to ensure that egg clutches are a result of a monogamous mating, *H. axyridis* should be separated at all stages. Newly emerged larva are fragile, and the best removal strategy needs to be implemented. A torn piece of paper placed in the petri dish with newly hatched larva leads to safe removal. Newly hatched larva will crawl on the paper, which can then be transferred to individual petri dishes. This paper also serves as a hiding place for newly emerged larva. Larva can feed on aphids immediately after emergence, so several aphids should be placed in the dish with each larva. If survival of all the larva is not the main goal of rearing, two or three larva can be placed together in one dish. This will result in sibling cannibalism, which will increase the speed of development for the surviving larva.

Emergence from pupal casing

Following the larval stage, *H. axyridis* enter a pupal stage. Upon emergence from the pupal stage into the adult stage, *H. axyridis* are bright yellow and spotless. Their final coloration and spot number will appear within 24 hours.

If genomic and transcriptomic work is the primary goal of the research, the newly emerged adults should not be fed. This strategy will reduce aphid contamination in the genome or transcriptome. If the research requires live, healthy adults, emerged adults should be fed a large supply (at least 10) of *A.*

pisum immediately. A damp cotton ball should also be placed in the habitat to prevent death due to dehydration.

2.3 Establishment of mating colonies and documenting progeny phenotypes

Establishment of a mating colony

Female *H. axyridis* are known to store sperm and in nature the average number of fathers per clutch can be as high as 3.8 (Ueno, 1996). To ensure monogamous matings, once a mating pair has been chosen, the female should not be exposed to any other males. If the female is needed as part of a different mating pair, the female should be separated from all males and held until she lays exclusively trophic eggs. This should ensure that all stored sperm has been eliminated and true paternities will be known for future egg clutches.

If egg clutch paternity is uncertain, observing the offspring phenotype can indicate non-monogamous mating pairs. Clutches that resulted from multiple fathers will likely display multiple phenotypes (**Figure 2.5a**), while clutches resulting from a single father will have highly similar phenotypes (**Figure 2.5b**).

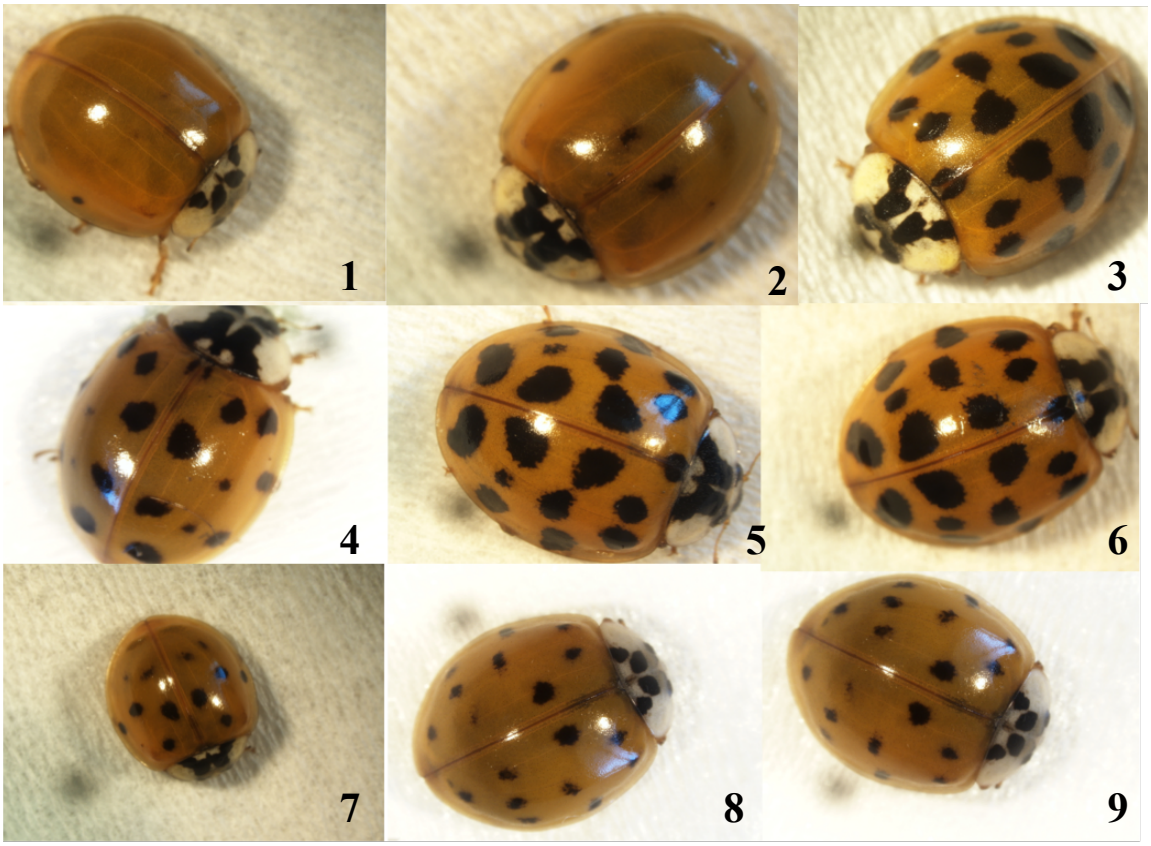

Figure 2.4a: A clutch of siblings that resulted from multiple males mating with a single female. Individuals 1 and 2 show a dramatically different spot number than individuals 3-9.

Figure 2.4b: A clutch of siblings that resulted from a single father mating with a single mother. Spot number phenotype is consistent across all individuals.

Documenting progeny phenotype

Documentation of individual phenotype is vital for tracing phenotypic patterns through the generations. In order to photograph and determine the gender of individual specimens, the *H. axyridis'* movement must first be slowed.

Individual *H. axyridis* were placed in 1.5 mL Eppendorf tubes (VWR ™). These tubes were placed either on ice or at 4° C. Cooling the specimens in a 4°C refrigerator is more efficient and reduces the time needed for cooling, thereby decreasing the likelihood of *H. axyridis* death. If using a 4ºC refrigerator, cooling for as little as ten minutes should be sufficient time for photographing.

A dissection microscope is ideal for recording the phenotype and sex of *H. axyridis*. The microscope should be enabled with a camera attachment to allow for multiple, independent verifications of specimen sex. Verifying sex by utilizing more than one trained individual is recommended to ensure mating pair success. In addition to sex determination, an in-laboratory microscope allows for the use of consistent settings to ensure accurate color representations throughout generations.

Figure 2.5: Sex determination of *Harmonia axyridis*. Sex was determined by looking at the second segment (circled above).

2.4 Summary

In summary, we have described how to establish an in-laboratory mating population of *H. axyridis*. Laboratory populations will enable researchers to track phenotypes over multiple generations. In addition, experiments can be conducted to tease apart environmental and genetic influences on phenotype. Collection of *H. axyridis* is easiest in the fall months, particularly October, but specimens can be collected during any season. Phenotypic characterization is simple, if care is taken to train individuals in sexing individuals.

To further establish *H. axyridis* as a model organism for the evolution of phenotypic variation, transcriptome and genome analyses are necessary. Chapter 3 will outline:

1. Development of a *de novo* transcriptome consisting of both an adult and larva life stage
2. Gene ontology analysis
3. Exploration of genetic differences between the life stages
4. Outline of the protein families present in adult and larval assemblies

CHAPTER III- EXPLORATION OF THE TRANSCIPTOME OF THE MULTICOLORED ASIAN LADY BEETLE, *HARMONIA AXYRIDIS*

Lindsay Havens [1*] and Matthew MacManes [1+]

1. University of New Hampshire Department of Molecular Cellular and Biomedical Science

* lao95@wildcats.unh.edu

+ matthew.macmanes@unh.edu

Code available here (https://goo.gl/2dnB8O).

3.1 Introduction:

Abstract:

The reasons for the evolution and maintenance of striking visual phenotypes are as widespread as the species that display these phenotypes. While study systems such as *Heliconius* and *Dendrobatidae* have been well characterized and provide critical information about the evolution of these traits, a breadth of new study systems, in which the phenotype of interest can be easily manipulated and quantified, are essential for gaining a more general understanding of these specific evolutionary processes. One such model is the multicolored Asian lady beetle, *Harmonia axyridis*, which displays significant elytral spot and color polymorphism. Using transcriptome data from two life

stages, adult and larva, we characterize the transcriptome, thereby laying a foundation for further analysis and identification of the genes responsible for the continual maintenance of spot variation in *H. axyridis*.

Introduction:

The evolution and maintenance of phenotypic polymorphism and striking visual phenotypes have fascinated scientists for many years (Darwin, 1859; Endler, 1986; Fisher, 1930; Gray and McKinnon, 2006; Joron et al., 2006). In general, insects have become increasingly popular as study organisms to examine phenotypic variation (Jennings, 2011; Joron et al., 2006). One such insect displaying extensive elytra and spot variation that has yet to be extensively studied is the Asian Multicolored Ladybeetle, *Harmonia axyridis*.

The mechanisms responsible for the evolution of these phenotypes are as widespread as the species that display them. Aposematism, crypsis, and mimicry may play a role in the evolution of phenotypic variation in the animal kingdom. Members of family *Dendrobatidae,* poison dart frogs, are aposematically colored (Cadwell, 1996), while *Tetrix subulata* grasshoppers maintain their phenotypic polymorphism to aid in crypsis (Karpestam et al., 2014). A mimicry strategy is utilized by one particularly well-characterized species that exhibits phenotypic polymorphism, the Neotropic butterfly genus, *Heliconius*. The color, pattern, and eyespot polymorphism seen in *Heliconius* is thought to have arose as a result of Müllerian mimicry (Flanagan et al., 2004) and the supergenes

underlying these traits have been well characterized (Kronforst et al., 2006, Joron et al., 2006, Jones et al., 2012).

These studies, aiming to elucidate the mechanistic links between phenotype and genotype, present a unique opportunity to gain insight into the inner workings of many important evolutionary processes. While systems like poison frogs and butterflies have been pioneering, the use of novel models, especially those that can be easily manipulated, are needed. One such study system that possesses many of the benefits of classical models, while offering several key benefits is the multicolored Asian lady beetle (=ALB), *Harmonia axyridis*. *Harmonia*, which is common throughout North America, and easily bred in laboratory environments, possesses significant variation in elytral spot number and color.

Elytra color can be red, orange, yellow, or black and spot numbers of *H. axyridis* range from zero to twenty-two (personal observation). The patterning is symmetrical on both wings. In some animals, there is a center spot beneath the pronotum which leads to an odd number of spots. The elytral spots are formed by the production of melanin pigments (Bezzerides et al., 2007). The frequency of different morphs varies with location and temperature. The melanic morph is more prevalent in Asia when compared to North America (LaMana and Miller, 1996, Dobzhansky, 1933). A decrease in melanic *H. axyridis* has been shown to be correlated with an increase in average yearly temperatures in the Netherlands (Brakefield and de Jong, 2011).

Sexual selection may play a role in color variation in *H. axyridis.* Osawa and Nishida (1992) remarked that female *H. axyridis* might choose their mates based on melanin concentration. Their choice, however, has been shown to vary based on season and temperature. Non-melanic (red, orange, or yellow with any spot number) males have a higher frequency of mating in the spring-time, while melanic (black) males have an increased frequency of mating in the summer. While this has been shown for elytral color, no such findings have occurred for spot number. Although these spot patterns are believed to be related to predator avoidance, thermotolerance, or mate choice (Osawa and Nishida, 1992), the genetics underlying these patterns is currently unknown.

To begin to understand the genomics of elytral spot patterning, we sequenced the transcriptome of an third instar larva and adult ladybug. These results lay the groundwork for future study of the genomic architecture of pigment placement and development in *H. axyridis*.

3.2 Methods and Materials:

Specimen capture, RNA extraction, library prep and sequencing

One adult and one larval *H. axyridis* were captured on the University of New Hampshire campus in Durham, New Hampshire (**Figure 3.1**). The adult was orange with 18 spots. The insects were immediately placed in RNAlater and stored in a -80C freezer until RNA extraction was performed. The RNA from both

individuals was extracted following the TRIzol extraction protocol (Invitrogen, Carlsbad USA). The entire insect was used for the RNA extraction protocol. The quantity and quality of extracted RNA was analyzed using a Qubit (Life Technology, Carlsbad USA) as well as a Tapestation 2200 (Agilent technologies, Palo Alto USA) prior to library construction. Following verification, RNA libraries were constructed for both samples following the TruSeq stranded RNA prep kit (Illumina, San Diego USA). To allow multiple samples to be run in one lane, a unique adapter was added to each sample. These samples were then pooled in equimolar quantities. The libraries were then sent to the New York Genome Center (New York, USA) for sequencing on a single lane (125bp paired end) of the HiSeq 2500 sequencer.

Figure 3.1: The adult and larva *Harmonia axyridis* that were used for RNA extraction and transcriptome analysis.

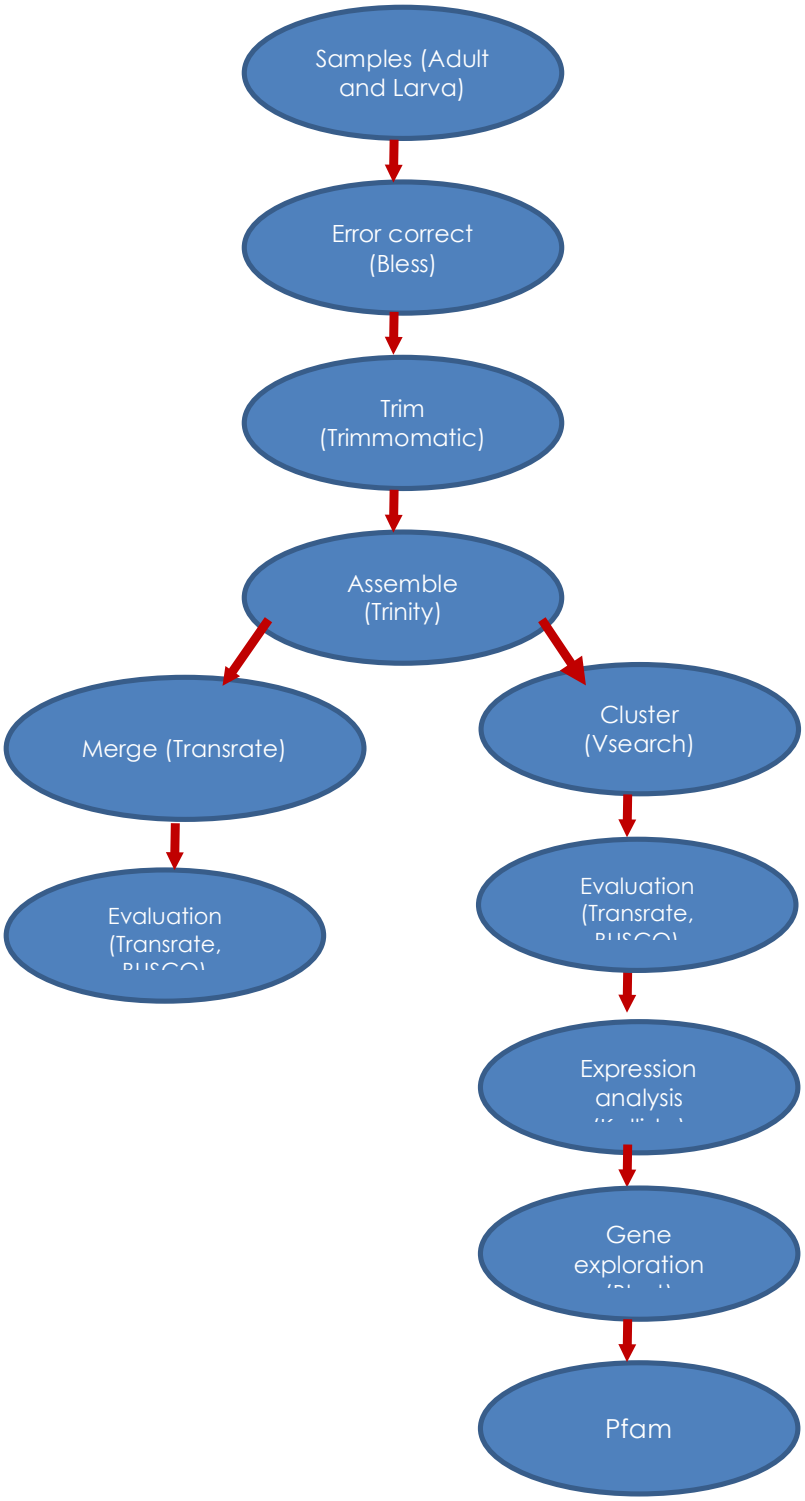

Figure 3.2: General pipeline followed for transcriptome building and analysis

Read processing and de novo assembly

Prior to assembly, reads from each individual were error corrected using BLESS (Heo et al., 2014) version 0.17 (https://goo.gl/tCF74w, https://goo.gl/nPLHtw). Following error correction, adapters were removed and reads were quality trimmed to a PHRED score of 2, using the software Trimmomatic (Boger, Lohse and Usadel, 2014), following recommendations outlined in MacManes and Eisen (2014). Together, these procedures resulted in the removal of less than 2% of reads. The remaining reads from each individual were assembled using independent executions of Trinity version 2.06 (Haas et al., 2013) and default settings (https://goo.gl/rPrlCr , https://goo.gl/rPrlCr).

Evaluating, optimizing, and merging adult and larva Trinity output files

The quality of the adult and larva assemblies was assessed via Transrate version 1.0 (Smith-Unna et al., 2015) using the BLESS error corrected reads. As part of this process, transrate used mapping metrics to filter out likely erroneously assembled transcripts, thereby producing an optimized (higher transrate scoring) assembly (https://goo.gl/0hLhN9). The initial transrate score for the assembly was 0.1539, while the optimized score was 0.2717. In addition to providing an optimized assembly, Transrate was also used to merge the adult and larva transcriptomes together, ultimately providing an improved transrate-merged assembly, which was used for downstream analyses.

As an alternative to the merging performed within Transrate, we used the clustering program VSEARCH version 3.0 (http://github.com/torognes/vsearch) to merge the adult and larva Trinity assemblies together. (https://goo.gl/tFwNuc). The resultant fasta file was analyzed via the Transrate pipeline as per above. The transrate assembly scores for each assembly were compared. The assembly with the highest score was retained for further analysis (https://goo.gl/0hLhN9)

Both the Transrate merged and the VSEARCH merged assemblies were analyzed with the BUSCO Version 1.1 pipeline (Simão et al., 2015) using both the Arthropoda and Eukaryota databases (https://goo.gl/0hLhN9 , https://goo.gl/tFwNuc). BUSCO provides an assessment of transcriptome and genome completeness by comparing genes present in experimental assembly with ubiquitous single-copy orthologs that are present in more than 90% of species in a given group.

To determine whether there was aphid transcript contamination in the final transcriptome assembly, aphid-base (www.aphidbase.com), an aphid genome database, was used for blastx analysis (States and Gish, 1994).

Assembled transcript exploration

To find relative abundance of each transcript, kallisto version 0.42.1 was used with default settings (Bray et al., 2015; https://goo.gl/9GSFvq). The 20 transcripts with the highest expression, based upon transcripts per million (TPM)

were selected from both the adult and larva transcriptomes (40 total). We also included the best 20 results from VSEARCH merged transcriptome. These were annotated with blastx using the UniProt database (www.uniprot.org) (https://goo.gl/qxgUZI). The blastx hit with the lowest e-value for each contig was identified. The UniProt identifier was then searched using Uniprot.org and the protein function was recorded.

Transcripts that were unique to larva and adult were analyzed using PANTHER Version 10.0 (pantherdb.org). To accomplish this, the UniProt identifiers (found using blastx) for these unique transcripts put into PANTHER, and gene lists and gene ontology pie charts were compiled.

To investigate gene ontology enrichment differences across the life stages of *H. axyridis*, AmiGO version 1.8 (http://amigo1.geneontology.org/cgi-bin/amigo/term_enrichment) was used. The UniProt identifiers for the unique transcripts in the adult and larva life stages were entered as gene products. The UniProt identifiers for transcripts found in both life stages were entered as background. The maximum p-value was set to .05.

TransDecoder was used to identify the complete open reading frames present in all assemblies. These data were then used to find Pfam protein families (Finn et al., 2014; https://goo.gl/SBWu3s, https://goo.gl/qe2mx3). An initial TransDecoder.LongOrfs run was used to determine the longest open reading frames present in each contig. The predicted ORFs were subsequently analyzed

with hmmer3 (Eddy, 2010) and blastp (Altschul et al., 1990). Hmmer3 performs a search for protein families using the Pfam database. Blastp provided a comprehensive list of proteins found within the transcriptomes using UniProt as a database. Using TransDecoder, the outputs of blastp and hmmer3 were used to predict peptides, coding sequences, and mRNA sequences. The resulting mRNA sequence was used to determine the number of complete open reading frames (https://goo.gl/qe2mx3).

3.3 Results and Discussion:

RNA extraction, assembly and evaluation

RNA was extracted from whole bodies of both the adult and the larva stage of a single *Harmonia axyridis*. The quality was verified using a Tapestation 2200 as well as a Qubit. The initial concentration for the larva sample was 83.2 ng/uL, while the initial concentration for the adult sample was 74.7 ng/uL. The number of strand-specific paired end reads contained in the adult and larva libraries were 58 million and 67 million, respectively. The reads were 125 base pairs in length. The raw reads will be made available in ENA and assigned a succession number when this manuscript is submitted.

The initial Trinity assemblies (will be made available on Dryad upon manuscript submission) were evaluated using TransDecoder (https://goo.gl/SBWu3s). Adult had more overall contigs and open reading frames than the larva life stage (**Table 3.1**). The adult transcriptome had 81,986

contigs while the larva had 68,876. Within those, the adult transcriptome had 19,235 (23.5%) complete open reading frames compared to 15,870 (19.8%) present in the larval transcriptome.

Table 3.1: Number of contigs and open reading frames from raw Trinity assemblies and improved Vsearch merged assembly

Trinity Assembly	Total Contigs	Number of complete open reading frames
Adult	114,726	19,849
Larva	96,586	15,870
Vsearch merged	154,473	27,088

Because we were interested in comparing the performance of two relatively new clustering algorithms, the adult and larva assemblies were merged with Transrate and VSEARCH (**Table 3.2**). Both merged files were filtered through Transrate, giving a Transrate-corrected assembly.

For both assemblies, the Transrate scores exceeded the threshold of 0.25, (indicative of high quality), and the BUSCO analyses indicated that the assemblies were highly complete in analyses conducted in the Arthropoda database (**Table 3.2**). In addition to BUSCO and Transrate scores, the percentage of good contigs was also taken into consideration, as well as the mean percent of open reading frames. Taken together, these data point to VSEARCH being slightly better than the Transrate merged assembly and the optimized VSEARCH assembly was used in downstream analysis.

As an additional quality control check, blastx was performed against the aphid-base database. There were no aphid specific hits, leading us to believe that if contamination was present, it was at very low levels.

Table 3.2: Data provided by Transrate and BUSCO for the VSEARCH and Transrate merged assemblies.

	Transrate				BUSCO % complete
Program used	**Transrate score**	**Optimal score**	**% Good contigs**	**Mean orf %**	**Arthropoda**
Transrate	0.1212	0.2827	67	56%	81%
VSEARCH	0.206	0.324	73	57%	81%

Gene exploration

The corrected reads and reference assembly, created by a combination of VSEARCH and Transrate described above, were processed by the kallisto pipeline (Bray et al., 2015). Three different kallisto analyses were performed. The first used the adult strand-specific, error corrected reads, the second used the larva strand-specific, error corrected reads, and the final used concatenated strand-specific, error corrected reads from the adult and larva combined transcripts.

The majority of transcripts had low expression (**Figures 3.3-3.5**), but the high expression transcripts were investigated further. The top 20 unique transcripts (meaning present in one life stage, but not the other) based on transcripts per million (TPM) were selected for both adult and larva. These contigs were then filtered through the blastx protocol using UniProt sequences as a database. The top hit for each unique contig was selected, and the protein corresponding to the Uniprot ID was recorded in **Tables 3.3, 3.4** & **3.5** (https://unh.box.com/s/g73xkz9mjm2vjgirwok9hxmmj52twklf) .

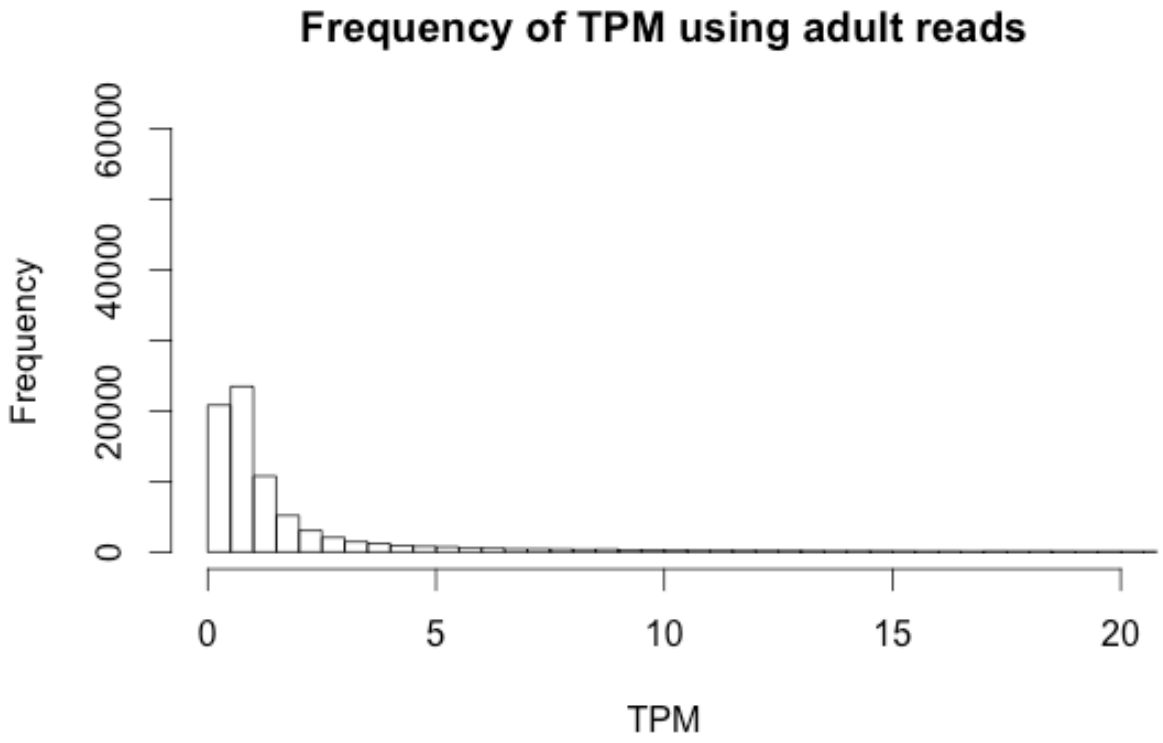

Figure 3.3: Frequency of TPM from kallisto using adult reads.

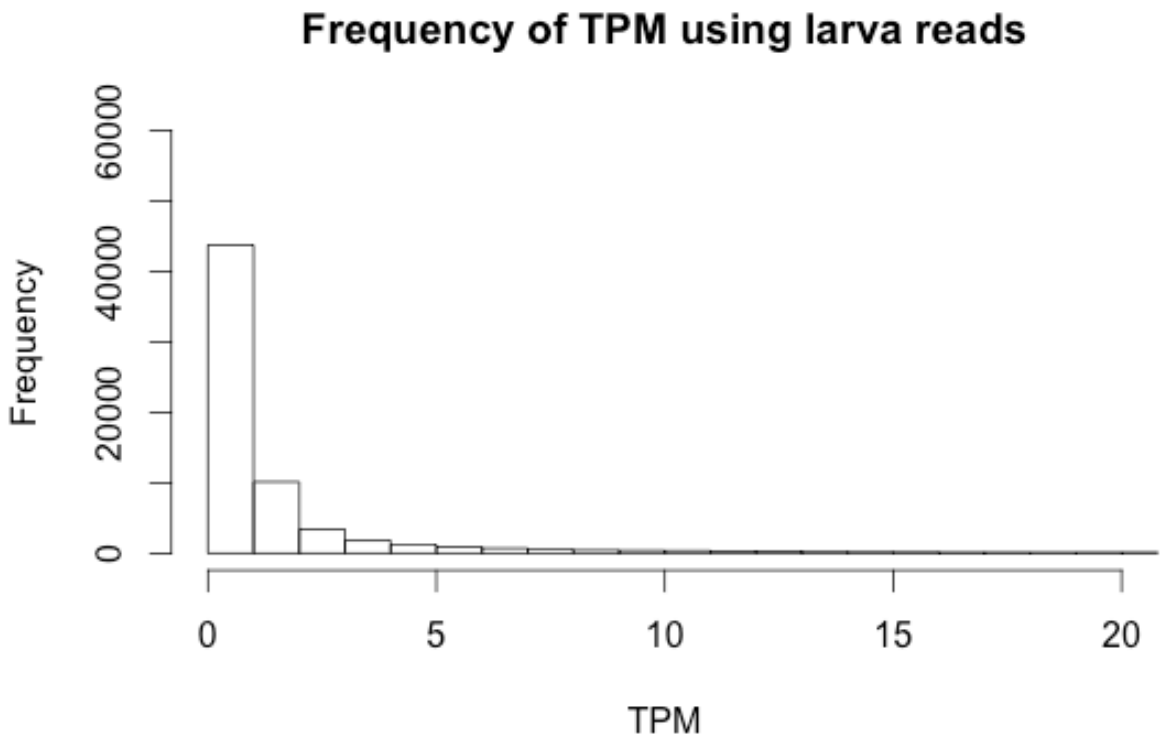

Figure 3.4: Frequency of TPM from kallisto using larva reads.

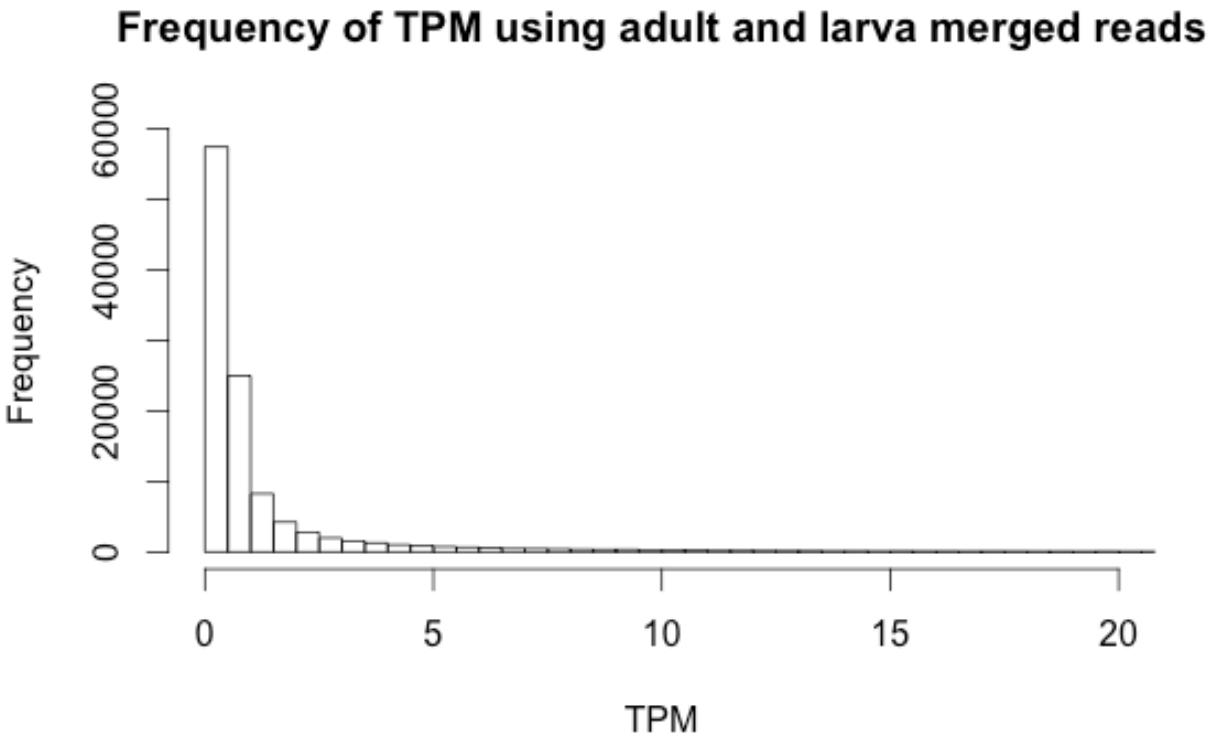

Figure 3.5: Frequency of TPM from kallisto using merged adult and larva transcripts.

Table 3.3: Most accurate unique contigs in the Vsearch merged transcriptome to which adult transcripts mapped. These were found using percent identity and e-value.

Transcript	Uniprot ID	% ID	Protein	Function
44492	sp | P79103 | RS4_BOVIN	97.34	40S ribosomal protein S4	poly(A) RNA binding
2272	sp | B0WWP2 | KLHDB_CULQU	91.58	Kelch-like protein diablo	Mediates ubiquitination and degredation of targe proteins
27163	sp | Q9VLN1 | WDR82_DROME	90.48	WD repeat-containing protein 82	Di and trimethylation of 'Lys-4' of histone H3
22204	sp | Q27268 | DX39B_DROME	90.02	ATP-dependent RNA helicase WM6	mRNA export out of the nucleus
18337	sp | Q9VR59 | VIP1_DROME	88.92	Inositol hexakisphosphate and diphosphoinositol-pentakisphosphate kinase	Synthesizes diphosphate addition
10059	sp | Q03720 | SLO_DROME	88.22	Calcium-activated potassium channel slowpoke	Potassium channel
11585	sp | O76484 | CSK2A_SPOFR	88.03	Casein kinase II subunit alpha	Wnt signaling
19426	sp | Q9VVE5 | MSIR6_DROME	87.98	RNA-binding protein Musashi homolog Rbp6	Regulates the expression of target mRNAs at the translation level

7157	sp | P92208 | JNK_DROME	87.67	Stress-activated protein kinase JNK	Response to environmental stress. Immune response
13257	sp | Q8IRI6 | GTR1_DROME	86.6	Glucose transporter type 1	Transports glucose
446	sp | Q27677 | ACES_LEPDE	84.76	Acetylcholinesterase	Hydrolyzes choline released into the synapse
17673	sp | Q7TSL3 | FBX11_RAT	84.63	F-box only protein 11	Substrate recognition component of SCF E3 ubiquitin-protein ligase complex
6146	sp | P41776 | GNAI_HOMAM	83.94	Guanine nucleotide-binding protein G(i) subunit alpha	Modulators or transducers in various transmembrane signaling systems
2588	sp | P17970 | KCNAB_DROME	82.89	Potassium voltage-gated channel protein Shab	Mediates voltage-dependent potassium ion permeability of excitable membranes
4273	sp | Q9V9J3 | SRC42_DROME	82.85	Tyrosine-protein kinase Src42A	Correct eye morphogenesis
20607	sp | Q9VZF4 | FBXW7_DROME	82.3	F-box/WD repeat-containing protein 7	Substrate recognition component of SCF E3 ubiquitin-protein ligase complex

13822	sp | P11833 | TBB_PARLI	82.04	Tubulin beta chain	Microtubule development
2039	sp | P35220 | CTNA_DROME	82	Catenin alpha	Associates with the cytoplasmic domain of cadherins
16320	sp | A2VEI2 | MICU1_DROME	81.97	Calcium uptake protein 1 homolog, mitochondrial	Key regulator of mitochondrial calcium uniporter

Table 3.4: Most accurate unique contigs in the Vsearch merged transcriptome to which larva transcripts mapped. These were found using percent identity and e-value.

Transcript ID	Uniprot ID	% ID	Protein	Function
26916	sp | Q8VDN2 | AT1A1_MOUSE	99.21	Sodium/potassium-transporting ATPase subunit alpha-1	Catalyzes hydrolysis of ATP
28360	sp | P32755 | HPPD_RAT	96.92	4-hydroxyphenylpyruvate dioxygenase	Degradation of tyrosine
41776	sp | Q03248 | BUP1_RAT	96.9	Beta-ureidopropionase	Amino-acid biosynthesis
12987	sp | Q06496 | NPT2A_RAT	96.26	Sodium-dependent phosphate transport protein 2A	Transporting phosphate
28453	sp | Q9DCU0 | MPRG_MOUSE	96.17	Membrane progestin receptor gamma	Steroid membrane receptor
45460	sp | P51655 | GPC4_MOUSE	95.54	Glypican-4	Development of kidney tubules and central nervous system
32850	sp | Q9QXN5 | MIOX_MOUSE	94.74	Inositol oxygenase	Myo-inositol + O2 = D-glucuronate + H2O.
24794	sp | P45952 | ACADM_MOUSE	94.54	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	Mitochondrial fatty acid beta-oxidation
22507	sp | Q17IE8 | CDK8_AEDAE	90.57	Cyclin-dependent kinase 8	Gene transciption of nearly all RNA polymerase II-dependent genes

27557	sp|Q9VLN1|WDR82_DROME	90.48	WD repeat-containing protein 82	Histon H3-K4 methylation
31029	sp|P37276|DYHC_DROME	89.64	Dynein heavy chain	Motor for the intracellular retrograde motility of vesicles
25254	sp|P07314|GGT1_RAT	89.52	Gamma-glutamyltranspeptidase 1	Cleaves the gamma-glutamyl bond of extracellular glutathione
11962	sp|P27590|UROM_RAT	87.2	Uromodulin	Biogenesis and organization of the apical membrane of epithelial cells
18585	sp|Q94900|GLUCL_DROME	87.14	Glutamate-gated chloride channel	Binding triggers a rapidly reversible current
39063	sp|Q8K023|AKC1H_MOUSE	87.04	Aldo-keto reductase family 1 member C18	Converts progesterone into 20 alpha-OHP
33883	sp|Q6ZQM8|UD17C_MOUSE	86.97	UDP-glucuronosyltransferase 1-7C	Conjunction and elmination of potentially toxic xenobiotics and endogenous compounds
6982	sp|P42124|EZ_DROME	86.31	Histone-lysine N-methyltransferase E(z)	Methylation of Lay-9 and Lys-27 of histone H3
5526	sp|P36872|2ABA_DROME	86.2	Protein phosphatase PP2A 55 kDa regulatory subunit	Substrate recognition function or targeting enzyme complexes to appropriate compartments
3906	sp|Q9VE46|SC5A7_DROME	85.54	High-affinity choline transporter 1	Imports Choline
26155	sp|P31430	85.3	Dipepridase 1	Hydrolyses of

	| DPEP1_RAT			dipeptides

Table 3.5: Most accurate contigs in the Vsearch merged transcriptome to which both the adult and larva transcripts mapped. These were found using percent identity and e-value.

Transcript ID	Uniprot ID	% ID	E-value	Protein	Function
11470	sp | P63018 | HSP7C_RAT	100	0	Heat shock cognate 71 kDa protein	Repressor of transcriptional activation
43677	sp | P0CG68 | UBC_PIG	100	0	Polyubiquitin -C	Activation of protein kinases and signaling
45310	sp | P63269 | ACTH_RAT	100	0	Actin, gamma-enteric smooth muscle	Cell motility
48128	sp | P26443 | DHE3_MOUSE	100	0	Glutamate dehydrogenase 1, mitochondrial	glutamine anaplerosis
43204	sp | P62919 | RL8_RAT	100	3.00E-177	60S ribosomal protein L8	Poly(A) RNA binding
34919	sp | P07823 | GRP78_MESAU	100	2.00E-172	78 kDa glucose-regulated protein	Assembly of multimeric protein complexes
42042	sp | P62425 | RL7A_RAT	100	1.00E-162	60S ribosomal protein L7a	Ribosome biogenesis
48088	sp | P06603 | TBA1_DROME	100	1.00E-158	Tubulin alpha-1 chain	Major constitute of microtubules
42846	sp | Q3SYU2 | EF2_BOVIN	100	2.00E-157	Elongation factor 2	Catalyzes the GTP-dependent ribosomal translocation step during translation elongation

54964	sp|Q9CQC8|SPG21_MOUSE	100	6.00E-151	Spg21	Negative regulation of T-cell activation
44638	sp|P62755|RS6_RAT	100	1.00E-150	40S ribosomal protein S6	Control of cell growth and proliferation
50383	sp|P62243|RS8_RAT	100	3.00E-150	40S ribosomal protein S8	Poly(A) RNA binding
48585	sp|P16125|LDHB_MOUSE	100	2.00E-149	L-lactate dehydrogenase B chain	(S)-lactate + NAD+ = pyruvate + NADH.
37462	sp|P62630|EF1A1_RAT	100	2.00E-139	Elongation factor 1-alpha 1	Promote binding of aminoacyl-tRNA to the A-site
55818	sp|P26284|ODPA_RAT	100	3.00E-137	Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	Links the glycolytic pathway to the tricarboxylic cycle
61840	sp|P09055|ITB1_MOUSE	100	1.00E-130	Integrin beta-1	Collagen receptors
65213	sp|P48721|GRP75_RAT	100	1.00E-124	Stress-70 protein, mitochondrial	Control of cell proliferation and aging
63181	sp|P12023|A4_MOUSE	100	5.00E-123	Amyloid beta A4 protein	Cell surface receptor
66438	sp|Q5I085|PP1B_XENTR	100	5.00E-119	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	Dephosphorylation of biological targets

57507	sp|Q8JFP1|IF4A2_CHICK	100	6.00E-114	Eukaryotic initiation factor 4A-II	Cap recognition

Most of the results were ubiquitous structural proteins. This indicates continuity in cellular structural processes throughout the life stages of *Harmonia axyridis*.

The blastx results from kallisto (adult and larva) runs were then put through the PANTHER (www.pantherdb.org) pipeline. The pie charts for molecular function and biological gene ontology can be seen below (**Figure 3.6**). These results indicate that both life stages have similar gene expression profiles.

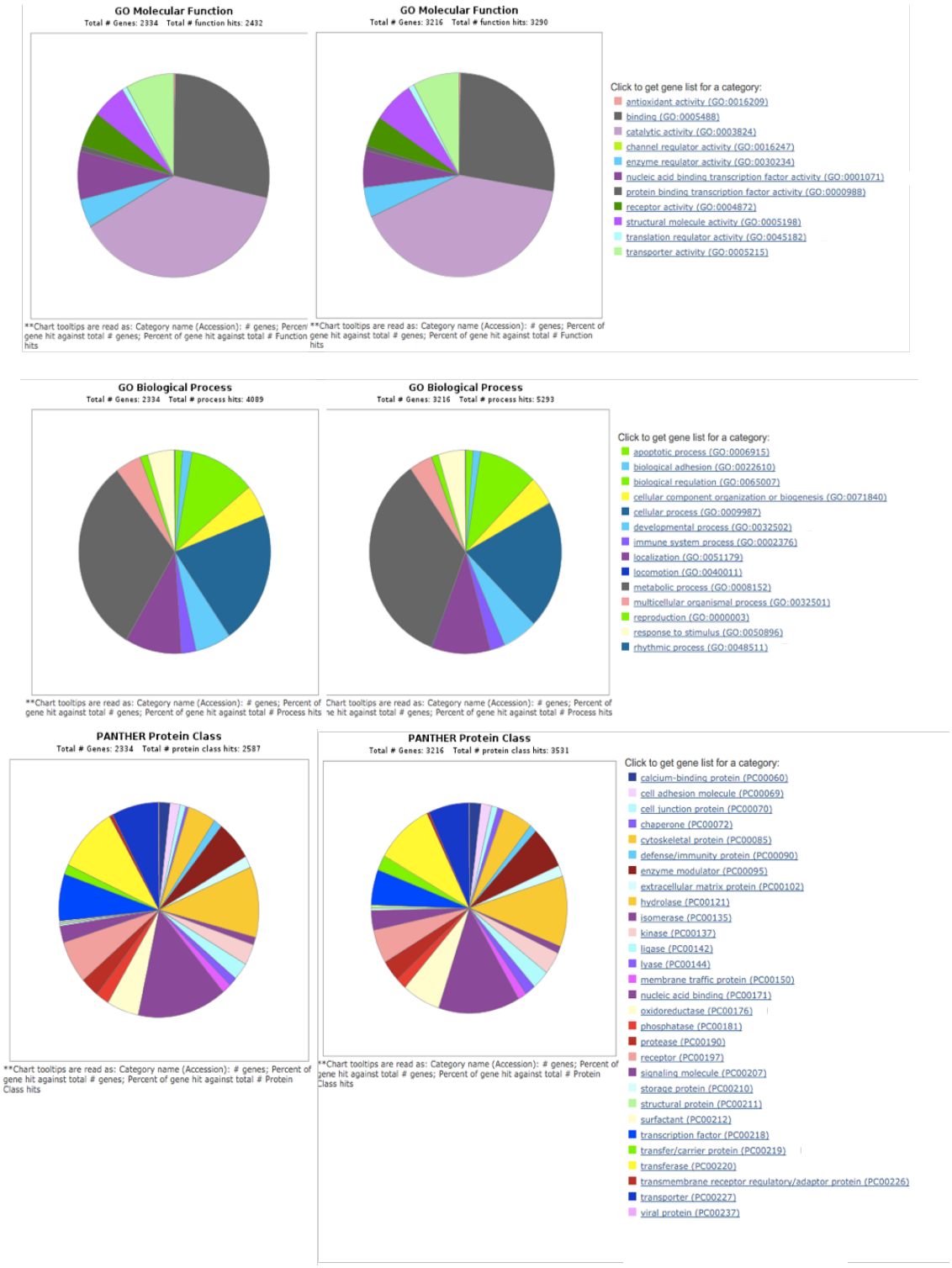

Figure 3.6: A comparison of unique PANTHER results from adult (left) and larva (right).

To explore the differences between the adult and larval life stages, gene ontology enrichment tests were performed using AmiGO version 1.8 (Ashburner et al., 2000) (**Figure 3.7**). These results indicate that in the adult life stage, the gene ontology profiles that are enriched are mainly synaptic and signaling proteins. No GO terms were enriched in the larva stage.

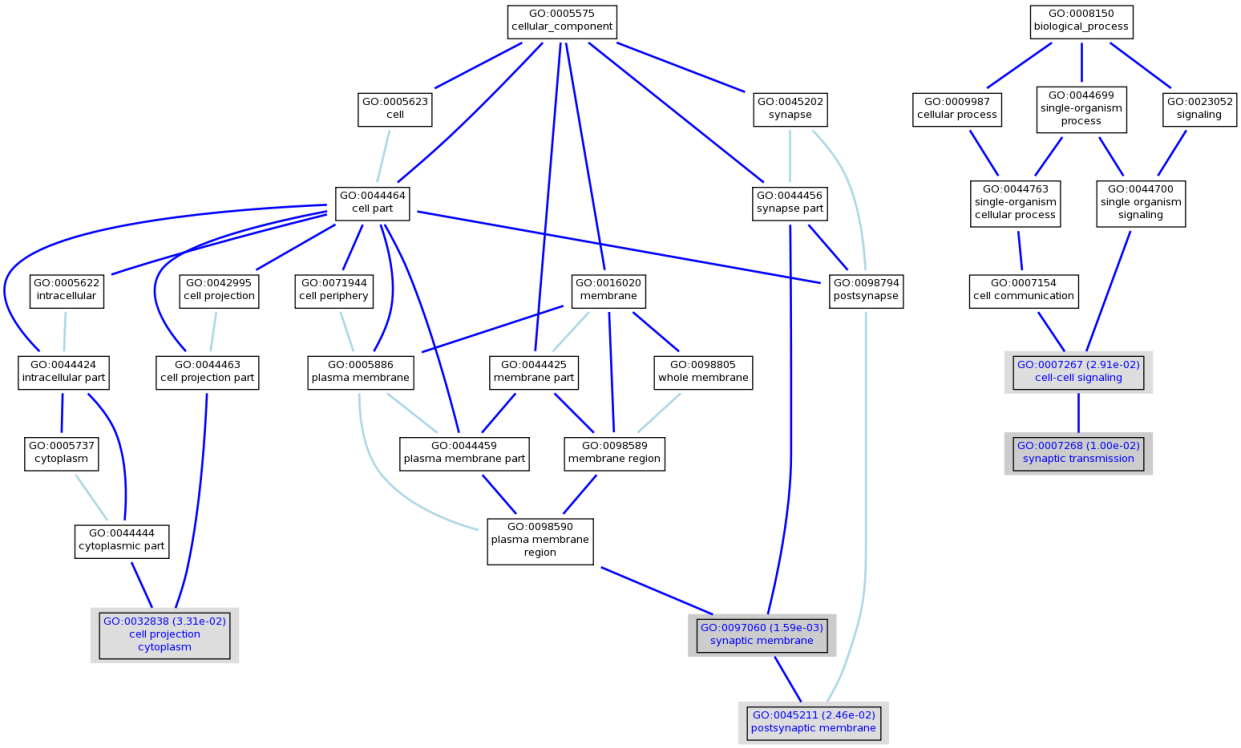

Figure 3.7: Gene ontology enrichment profile of the adult stage *H. axyridis*.

***Estimation of transcript expression**:*

Gene expression was estimated using number of transcripts present in the kallisto data. The results indicate 76,920 transcripts are ubiquitous across life stages and between individuals (**Figure 3.8**). This analysis also indicates that the adult life contains more unique genes (44,048) when compared to the larval life stage (30,903). The proteins that were unique to certain life stages were present in one stage, while completely absent in the other.

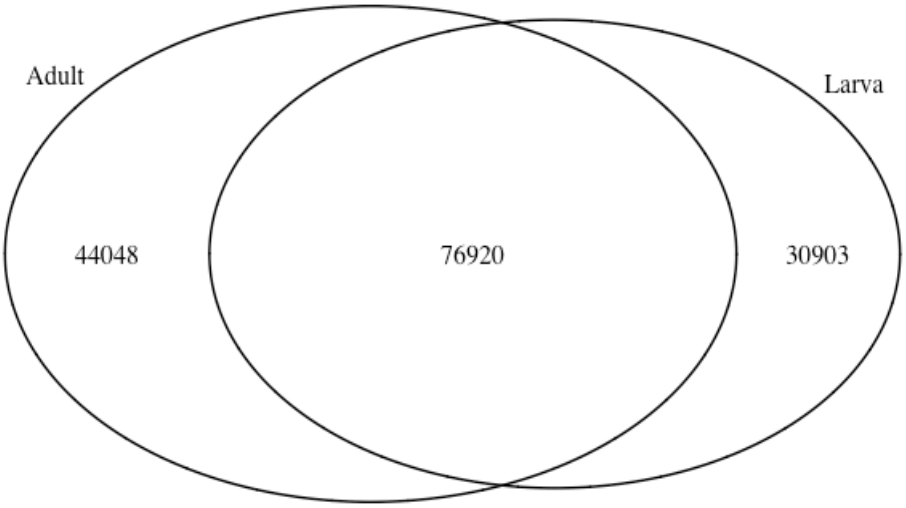

Figure 3.8: A Venn diagram for gene expression of transcripts in the adult and larva stages.

TransDecoder and Pfam analysis-adult transcriptome

TransDecoder was used to find a comprehensive list of the longest open reading frames from each contig present in the trinity-assembled adult transcriptome. These were then used to identify protein families using the hmmer3 pipeline previously described. These results, as well as blastp results, were used to find complete open reading frames. For the adult transcriptome, the number of complete open reading frames was 19,849 (**Table 3.1**).

Complete Pfam results were filtered using Pfam hit e-value. Although the top 200 protein families were analyzed, the top 50 Pfam unique matches for the transcriptome are shown (https://unh.box.com/s/bxrx8o8wjmi5br0e2cu6bi5mauevi6rp) . These 50 proteins were mostly structural in function (**Table 3.6**), though there were two heat shock proteins identified, HSP70 and HSP90. Further analysis of additional Pfam proteins outside this filtered top 50 data set identified one pigment protein family, retinal pigment epithelial membrane proteins, and was included in **Table 3.6**. However, the E-value for this match was higher than those for matches in the best 50.

Table 3.6: The top 50 Pfam family hits found using the adult trinity-assembled transcriptome

Target	Accession	Query	E - value	Protein Family	Function
Aquarius_N	PF16399.1	TRINITY_DN31573_c0_g1_i2|m.15760	0	Intron-binding protein aquarius N-terminus	Splicing factor
Cadherin	PF00028.13	TRINITY_DN35646_c0_g1_i1|m.32481	0	Cadherin domain	Cell adhesion
E3_UbLigase_R4	PF13764.2	TRINITY_DN35614_c1_g1_i2|m.32427	0	E3 ubiquitin-protein ligase UBR4	Ubiquitin ligase enzyme
EGF_CA	PF07645.11	TRINITY_DN35574_c0_g1_i1|m.32344	0	Calcium-binding EGF domain	Membrane-bound proteins
fn3	PF00041.17	TRINITY_DN35636_c0_g1_i2|m.32539	0	Fibronectin type III domain	Extracellular proteins
FragX_IP	PF05994.7	TRINITY_DN23915_c0_g1_i1|m.4721	0	Cytoplasmic Fragile-X interacting family	Unknown
Glycogen_syn	PF05693.9	TRINITY_DN29821_c0_g1_i1|m.11579	0	Glycogen synthase	Synthesis of the -1,4 linked glucose chain
I-set	PF07679.12	TRINITY_DN35642_c0_g1_i1|m.32416	0	Immunoglobulin I-set domain	Cell adhesion, hemolymph, and structural proteins
INTS2	PF14750.2	TRINITY_DN32471_c0_g1_i1|m.18245	0	Integrator complex subunit 2	snRNA transcription and processing
Ldl_recept_a	PF00057.14	TRINITY_DN35560_c0_g1_i1|m.32304	0	Low-density lipoprotein receptor domain class A	Cell surface receptors
Med23	PF11573.4	TRINITY_DN32046_c0_g1_i1|m.17180	0	Mediator complex subunit 23	Transcription activation
Myosin_tail_1	PF01576.15	TRINITY_DN16526_c0_g1_i1|m.2165	0	Myosin tail	Myosin
Nup192	PF11894.4	TRINITY_DN34901_c0_g1_i1|m.28630	0	Nuclear pore complex scaffold, nucleoporins 186/192/205	Uncharacterized

Phosphorylase	PF00343.16	TRINITY_DN34425_c2_g1_i1|m.26321	0	Carbohydrate phosphorylase	Forms glucose-1 phosphate
Spectrin	PF00435.17	TRINITY_DN3797_c0_g1_i1|m.371	0	Spectrin repeat	Cytoskeleton structure
Strumpellin	PF10266.5	TRINITY_DN31874_c0_g1_i1|m.16647	0	Hereditary spastic paraplegia protein strumpellin	Structural platform for cytoskeleton protein assembly
UDP-g_GGTase	PF06427.7	TRINITY_DN31001_c0_g1_i1|m.14440	0	UDP-glucose:Glycoprotein Glucosyltransferase	Verifies glycoproteins are folded correctly
V_ATPase_I	PF01496.15	TRINITY_DN34359_c2_g1_i1|m.25809	0	V-type ATPase 116kDa subunit family	Acidifies intracellular organelles, proton pump
Vinculin	PF01044.15	TRINITY_DN34029_c0_g1_i1|m.24498	0	Vinculin family	Attachment of actin-based microfiliments to plasma membrane
Mon2_C	PF16206.1	TRINITY_DN21615_c0_g4_i1|m.3622	4.00E-304	C-terminal region of Mon2 protein	Endo membrane trafficking
Myosin_head	PF00063.17	TRINITY_DN34376_c2_g2_i1|m.25676	2.10E-299	Myosin head (motor domain)	Myosin
CUB	PF00431.16	TRINITY_DN35593_c1_g1_i1|m.32241	1.80E-297	CUB domain	Conserved protein domain
An_peroxidase	PF03098.11	TRINITY_DN35209_c0_g1_i1|m.30336	5.50E-294	Animal haem peroxidase	Oxidative stress defense
Filamin	PF00630.15	TRINITY_DN33211_c0_g1_i2|m.21405	4.70E-288	Filamin/ABP2 80 repeat	Unknown
TH1	PF04858.9	TRINITY_DN17207_c0_g1_i1|m.2306	1.00E-287	TH1 protein	Uncharacterized
ST7	PF04184.8	TRINITY_DN32837_c0_g1_i1|m.19710	4.50E-286	ST7 protein	Tumor suppressor
Nckap1	PF09735.5	TRINITY_DN34037_c0_g1_i1|m.24500	1.80E-285	Membrane-associated apoptosis protein	Acting polymerization; WAVE2 signaling complex
Vps35	PF03635.13	TRINITY_DN28501_c0_g1_i2|m.9409	2.90E-284	Vacuolar protein sorting-associated protein 35	Protein Trafficking
Dynein_heavy	PF03028.11	TRINITY_DN24311_c0_g1_i1|m.4968	7.90E-277	Dynein heavy chain and region D6 of	ATP chemical energy to mechanical

					energy
PEPCK	PF00821.14	TRINITY_DN32319_c0_g2_i1|m.17968	2.40E-271	Phosphoenolpyruvate carboxykinase	Gluconeogenesis pathway
HSP70	PF00012.16	TRINITY_DN29127_c1_g1_i1|m.10267	1.50E-269	Hsp70 protein	Protein Folding
Med24_N	PF11277.4	TRINITY_DN30181_c0_g1_i1|m.12363	2.60E-268	Mediator complex subunit 24 N-terminal	Gene expression
Dymeclin	PF09742.5	TRINITY_DN21165_c0_g1_i2|m.3387	5.50E-267	Dyggve-Melchior-Clausen syndrome protein	Vesicle secretion or inter-cellular signaling
ACC_central	PF08326.8	TRINITY_DN34812_c2_g1_i1|m.28364	6.20E-265	Acetyl-CoA carboxylase, central region	Synthesis of long-chain fatty acids
FSA_C	PF10479.5	TRINITY_DN35631_c0_g1_i1|m.32465	2.40E-261	Fragile site-associated protein C-terminus	C-termius half of KIAA1109l
Glyco_transf_41	PF13844.2	TRINITY_DN28084_c0_g1_i1|m.8708	1.80E-256	Glycosyl transferase family 41	Ser and Thr formation
Plexin_cytopl	PF08337.8	TRINITY_DN24621_c0_g1_i1|m.5182	1.50E-255	Plexin cytoplasmic RasGAP domain	Semaphoring receptors
DUF2151	PF10221.5	TRINITY_DN29166_c0_g1_i1|m.10343	2.60E-253	Cell cycle and development regulator	S-M cycles
UNC-79	PF14776.2	TRINITY_DN33919_c0_g1_i1|m.23909	1.10E-252	Cation-channel complex subunit UNC-79	Cation-channel complex
Hid1	PF12722.3	TRINITY_DN21165_c0_g1_i2|m.3387	1.20E-251	High-temperature-induced dauer-formation protein	Vesicle secretion or inter-cellular signaling
FTHFS	PF01268.15	TRINITY_DN32053_c1_g1_i1|m.17291	3.30E-248	Formate--tetrahydrofolate ligase	Catalytic enzyme
PGI	PF00342.15	TRINITY_DN30271_c0_g1_i1|m.12459	2.10E-243	Phosphoglucose isomerase	Fructose 6-phosphate formation
Menin	PF05053.9	TRINITY_DN33385_c0_g1_i1|m.21772	1.10E-241	Menin	Tumor suppressor gene
Ion_trans	PF00520.27	TRINITY_DN34557_	7.10E-	Ion transport	Ion transport

		c0_g1_i2|m.26858	241	protein	
HSP90	PF00183.14	TRINITY_DN22165_ c0_g1_i1|m.3857	8.50E-240	Hsp90 protein	Protein folding and degradation
KAP	PF05804.8	TRINITY_DN12017_ c0_g1_i1|m.1330	6.00E-239	Kinesin-associated protein (KAP)	Protein folding
Strabismus	PF06638.7	TRINITY_DN34132_ c0_g1_i1|m.24946	1.80E-238	Strabismus protein	Unknown
PROCN	PF08083.7	TRINITY_DN35422_ c1_g1_i1|m.31616	6.00E-234	PROCN (NUC071) domain	Pre-mRNA splicing
GDI	PF00996.14	TRINITY_DN32537_ c1_g1_i1|m.18703	3.00E-232	GDP dissociation inhibitor	Regulate vesicular membrane trafficking
RPE65	PF03055.11	TRINITY_DN33976_ c4_g1_i1|m.24265	2.90E-102	Retinal pigment epithelial membrane protein	Cleavage of carotenoids

TransDecoder and Pfam analysis- larva transcriptome

As described in the previous section, Pfam proteins and complete open reading frames were found. The total number of complete open reading frames was found to be 15,870 (**Table 3.1**).

Again, most of the protein families found were structural proteins. Analysis of the of the best Pfam matches (**Table 3.7**) indicated that there were unique protein families in larva stage when compared to the adult life stage. Additionally, in the larva transcriptome, one pigment and one melanin protein family was found and included at the bottom of **Table 3.7).** Again, these E-values were lower than the best 50 matches.

Table 3.7: The top 50 Pfam family hits found using the larva trinity-assembled transcriptome

Target	Accession	Query	E-value	Protein Family	Function
E3_UbLigase_R4	PF13764.2	TRINITY_DN30245_c0_g1_i1|m.30408	0	E3 ubiquitin-protein ligase UBR4	Ubiquitin ligase enzyme
fn3	PF00041.17	TRINITY_DN21160_c0_g1_i2|m.7309	0	Fibronectin type III domain	Extracellular proteins
FragX_IP	PF05994.7	TRINITY_DN30436_c0_g1_i1|m.31270	0	Cytoplasmic Fragile-X interacting family	Unknown
Glycogen_syn	PF05693.9	TRINITY_DN25451_c1_g1_i1|m.13959	0	Glycogen synthase	Synthesis of the -1,4 linked glucose chain
I-set	PF07679.12	TRINITY_DN21160_c0_g1_i2|m.7309	0	Immunoglobulin I-set domain	Cell adhesion, hemolymph, and structural proteins
Med23	PF11573.4	TRINITY_DN29779_c0_g1_i1|m.28322	0	Mediator complex subunit 23	Transcription activation
Myosin_head	PF00063.17	TRINITY_DN27752_c0_g1_i2|m.20422	0	Myosin head (motor domain)	Myosin
Myosin_tail_1	PF01576.15	TRINITY_DN26661_c0_g3_i1|m.16785	0	Myosin tail	Myosin
Nckap1	PF09735.5	TRINITY_DN29660_c0_g1_i1|m.27812	0	Membrane-associated apoptosis protein	Acting polymerization; WAVE2 signaling complex
Phosphorylase	PF00343.16	TRINITY_DN27496_c0_g1_i1|m.19191	0	Carbohydrate phosphorylase	Forms glucose-1 phosphate
Spectrin	PF00435.17	TRINITY_DN29520_c1_g1_i1|m.27316	0	Spectrin repeat	Cytoskeleton structure
Strumpellin	PF10266.5	TRINITY_DN28363_c0_g2_i3|m.22541	0	Hereditary spastic paraplegia protein strumpellin	Structural platform for cytoskeleton protein assembly

UDP-g_GGTase	PF06427.7	TRINITY_DN25814_c0_g1_i1|m.14697	0	UDP-glucose:Glycoprotein Glucosyltransferase	Verifies glycoproteins are folded correctly
V_ATPase_I	PF01496.15	TRINITY_DN26256_c1_g1_i1|m.15871	0	V-type ATPase 116kDa subunit family	Acidifies intracellular organelles, proton pump
Vinculin	PF01044.15	TRINITY_DN28966_c1_g1_i1|m.24942	0	Vinculin family	Attachment of actin-based microfiliments to plasma membrane
Mon2_C	PF16206.1	TRINITY_DN25083_c0_g1_i1|m.12981	2.50E-303	C-terminal region of Mon2 protein	Endo membrane trafficking
An_peroxidase	PF03098.11	TRINITY_DN30075_c1_g1_i1|m.29879	1.20E-293	Animal haem peroxidase	Oxidative stress defense
EGF_CA	PF07645.11	TRINITY_DN30293_c0_g1_i1|m.30832	4.10E-291	Calcium-binding EGF domain	Membrane-bound proteins
TH1	PF04858.9	TRINITY_DN25206_c0_g1_i1|m.13426	1.00E-287	TH1 protein	Uncharacterized
ST7	PF04184.8	TRINITY_DN27733_c0_g1_i1|m.20413	4.50E-286	ST7 protein	Tumor suppressor
Vps35	PF03635.13	TRINITY_DN9274_c0_g1_i1|m.1809	2.90E-284	Vacuolar protein sorting-associated protein 35	Protein Trafficking
INTS2	PF14750.2	TRINITY_DN25452_c0_g1_i2|m.13923	1.40E-275	Integrator complex subunit 2	snRNA transcription and processing
PEPCK	PF00821.14	TRINITY_DN26352_c0_g1_i1|m.16103	6.60E-271	Phosphoenol pyruvate carboxykinase	Gluconeogenesis pathway
HSP70	PF00012.16	TRINITY_DN11800_c0_g1_i1|m.2486	8.70E-271	Hsp70 protein	Protein folding
Med24_N	PF11277.4	TRINITY_DN25700_c0_g1_i1|m.14534	5.70E-268	Mediator complex subunit 24 N-terminal	Gene expression
Dymeclin	PF09742.5	TRINITY_DN29673_c0_g1_i1|m.27881	5.20E-267	Dyggve-Melchior-Clausen syndrome protein	Vesicle secretion or inter-cellular signaling
ACC_central	PF08326.8	TRINITY_DN29774_c1_g1_i1|m.28166	6.60E-265	Acetyl-CoA carboxylase, central region	Synthesis of long-chain fatty acids

FSA_C	PF10479.5	TRINITY_DN30339_c0_g1_i4|m.31084	1.40E-262	Fragile site-associated protein C-terminus	C-termius half of KIAA1109l
Glyco_transf_41	PF13844.2	TRINITY_DN24610_c0_g1_i1|m.12154	1.80E-256	Glycosyl transferase family 41	Ser and Thr formation
Plexin_cytopl	PF08337.8	TRINITY_DN10059_c0_g2_i1|m.1950	1.40E-255	Plexin cytoplasmic RasGAP domain	Semaphoring receptors
DUF2151	PF10221.5	TRINITY_DN27984_c0_g1_i1|m.20973	2.70E-253	Cell cycle and development regulator	S-M cycles
Hid1	PF12722.3	TRINITY_DN29673_c0_g1_i1|m.27881	8.50E-252	High-temperature-induced dauer-formation protein	Vesicle secretion or inter-cellular signaling
FTHFS	PF01268.15	TRINITY_DN25971_c1_g1_i2|m.14981	5.10E-248	Formate--tetrahydrofolate ligase	Catalytic enzyme
Cadherin	PF00028.13	TRINITY_DN30308_c0_g2_i2|m.31154	4.90E-247	Cadherin domain	Cell adhesion
HSP90	PF00183.14	TRINITY_DN18245_c0_g1_i1|m.4967	1.50E-244	Hsp90 protein	Protein folding and degradation
PGI	PF00342.15	TRINITY_DN26852_c0_g1_i1|m.17463	1.80E-243	Phosphoglucose isomerase	Fructose 6-phosphate formation
Dynein_heavy	PF03028.11	TRINITY_DN27832_c0_g1_i2|m.20601	2.00E-242	Dynein heavy chain and region D6 of	ATP chemical energy to mechanical energy
KAP	PF05804.8	TRINITY_DN27994_c1_g1_i2|m.21287	6.00E-239	Kinesin-associated protein (KAP)	Protein folding
Ion_trans	PF00520.27	TRINITY_DN28629_c0_g1_i1|m.23740	2.20E-235	Ion transport protein	Ion transport
PROCN	PF08083.7	TRINITY_DN30083_c0_g1_i1|m.29585	6.00E-234	PROCN (NUC071) domain	Pre-mRNA slicing factors
GDI	PF00996.14	TRINITY_DN22802_c1_g1_i1|m.9225	3.00E-232	GDP dissociation inhibitor	Regulate vesicular membrane trafficking
SNF	PF00209.14	TRINITY_DN30058_c0_g1_i1|m.29743	1.30E-229	Sodium:neurotransmitter symporter family	Release, re-uptake and recycling of neurotransmitters

eIF-3c_N	PF05470.8	TRINITY_DN24381_c1_g1_i1|m.11657	5.30E-229	Eukaryotic translation initiation factor 3 subunit 8	Early step of translation initiation
tRNA-synt_2c	PF01411.15	TRINITY_DN32258_c0_g1_i1|m.31545	1.10E-227	tRNA synthetases class II (A)	tRNA synthetase
Glyco_hydro_18	PF00704.24	TRINITY_DN30173_c0_g1_i1|m.30266	3.10E-227	Glycosyl hydrolases family 18	Hydrolyzes bond between carbohydrates
COBRA1	PF06209.9	TRINITY_DN28230_c0_g1_i1|m.22089	1.50E-226	Cofactor of BRCA1 (COBRA1)	Chromatin unfolding
Macoilin	PF09726.5	TRINITY_DN26432_c0_g2_i1|m.16195	3.60E-224	Transmembrane protein	Transmembrane proteins
HCO3_cotransp	PF00955.17	TRINITY_DN29948_c0_g1_i13|m.29021	6.60E-223	HCO3-transporter family	Bicarbonate transporter
Peptidase_M49	PF03571.11	TRINITY_DN23169_c0_g1_i1|m.9779	9.70E-222	Peptidase family M49	Hydrolysis of peptide bonds
HgmA	PF04209.9	TRINITY_DN28834_c0_g1_i1|m.24328	6.50E-221	homogentisate 1,2-dioxygenase	Cleaves the aromatic ring during degredation of Phe and Tyr
RPE65	PF03055.11	TRINITY_DN26925_c0_g1_i4|m.17653	2.20E-102	Retinal pigment epithelial membrane protein	Cleavage of carotenoids
Pro-MCH	PF05824.8	TRINITY_DN22889_c0_g3_i1|m.9278	0.057	Pro-melanin-concentrating hormone (Pro-MCH)	Stimulates hunger and may additionally regulate energy homeostasis, reproductive function, and sleep

TransDecoder and Pfam analysis- Vsearch transcriptome

Finally, the pipeline described above was utilized to find the complete open reading frames and most accurate Pfam matches for the Vsearch merged transcriptome. The number of complete open reading frames was 27,088 (**Table 3.1**).

Again, most of the protein families found were structural proteins (**Table 3.8**). Many of these proteins are unique when compared to the larva and adult transcriptomes. In addition, upon further analysis of the Pfam output, one pigment protein and one melanin protein family was discovered and are shown at the bottom of **Table 3.8**. The E-value for these protein families was low when compared to the rest of proteins present in **Table 3.8**.

Table 3.8: The top 50 Pfam family hits found using the Vsearch merged transcriptome

Target	Accession	Query	E-Value	Protein Family	Function
Aquarius_N	PF16399.1	2158|m.2982	0	Intron-binding protein aquarius N-terminus	Splicing factor
Cadherin	PF00028.13	10|m.7	0	Cadherin domain	Cell adhesion
E3_UbLigase_R4	PF13764.2	5|m.2	0	E3 ubiquitin-protein ligase UBR4	Ubiquitin ligase enzyme
EGF_CA	PF07645.11	118|m.227	0	Calcium-binding EGF domain	Membrane-bound proteins
FragX_IP	PF05994.7	2383|m.3246	0	Cytoplasmic Fragile-X interacting family	Unknown
Glycogen_syn	PF05693.9	6590|m.8036	0	Glycogen synthase	Synthesis of the -1,4 linked glucose chain
I-set	PF07679.12	8|m.4	0	Immunoglobulin I-set domain	Cell adhesion, hemolymph, and structural proteins
INTS2	PF14750.2	3747|m.4869	0	Integrator complex subunit 2	snRNA transcription and processing
Ldl_recept_a	PF00057.14	18|m.24	0	Low-density lipoprotein receptor domain class A	Cell surface receptors
Med23	PF11573.4	2381|m.3244	0	Mediator complex subunit 23	Transcription activation
Myosin_head	PF00063.17	473|m.758	0	Myosin head (motor domain)	Myosin
Myosin_tail_1	PF01576.15	328|m.560	0	Myosin tail	Myosin
Nckap1	PF09735.5	3891|m.5038	0	Membrane-associated apoptosis protein	Acting polymerization ; WAVE2 signaling complex
Nup192	PF11894.4	664|m.1026	0	Nuclear pore complex scaffold, nucleoporins 186/192/205	Uncharacterized
Phosphorylase	PF00343.16	6863|m.8356	0	Carbohydrate phosphorylase	Forms glucose-1 phosphate
Spectrin	PF00435.17	38|m.59	0	Spectrin repeat	Cytoskeleton structure
Strumpellin	PF10266.5	3691|m.4804	0	Hereditary spastic paraplegia protein strumpellin	Structural platform for cytoskeleton protein assembly

UDP-g_GGTase	PF06427.7	896 | m.1338	0	UDP-glucose:Glycoprotein Glucosyltransferase	Verifies glycoproteins are folded correctly
V_ATPase_I	PF01496.15	5813 | m.7202	0	V-type ATPase 116kDa subunit family	Acidifies intracellular organelles, proton pump
Vinculin	PF01044.15	2468 | m.3347	0	Vinculin family	Attachment of actin-based microfiliments to plasma membrane
Mon2_C	PF16206.1	3999 | m.5166	4.00E-304	C-terminal region of Mon2 protein	Endo membrane trafficking
CUB	PF00431.16	334 | m.572	1.80E-297	CUB domain	Conserved protein domain
An_peroxidase	PF03098.11	1557 | m.2225	5.50E-294	Animal haem peroxidase	Oxidative stress defense
Filamin	PF00630.15	244 | m.429	4.70E-288	Filamin/ABP280 repeat	Unknown
TH1	PF04858.9	14283 | m.15684	1.00E-287	TH1 protein	Uncharacterized
ST7	PF04184.8	15180 | m.16497	4.50E-286	ST7 protein	Tumor suppressor
Vps35	PF03635.13	7588 | m.9149	2.90E-284	Vacuolar protein sorting-associated protein 35	Protein Trafficking
Dynein_heavy	PF03028.11	10314 | m.11897	7.90E-277	Dynein heavy chain and region D6 of dynein	ATP chemical energy to mechanical energy
PEPCK	PF00821.14	9457 | m.11023	2.40E-271	Phosphoenolpyruvate carboxykinase	Gluconegenesis pathway
HSP70	PF00012.16	11470 | m.13021	8.70E-271	Hsp70 protein	Protein Folding
Med24_N	PF11277.4	5837 | m.7226	2.60E-268	Mediator complex subunit 24 N-terminal	Gene expression
Dymeclin	PF09742.5	5583 | m.6957	5.20E-267	Dyggve-Melchior-Clausen syndrome protein	Vesicle secretion or inter-cellular signaling
ACC_central	PF08326.8	263 | m.456	6.20E-265	Acetyl-CoA carboxylase, central region	Synthesis of long-chain fatty acids
Glyco_transf_41	PF13844.2	3093 | m.4094	1.80E-256	Glycosyl transferase family 41	Ser and Thr formation
Plexin_cytopl	PF08337.8	1306 | m.1895	1.40E-255	Plexin cytoplasmic RasGAP domain	Semaphoring receptors
DUF2151	PF10221.5	8981 | m.10557	2.60E-253	Cell cycle and development regulator	S-M cycles
UNC-79	PF14776.2	8243 | m.9821	1.10E-252	Cation-channel complex subunit	Cation-channel

				UNC-79	complex
Hid1	PF12722.3	5583 | m.6957	8.50E-252	High-temperature-induced dauer-formation protein	Vesicle secretion or inter-cellular signaling
FTHFS	PF01268.15	6062 | m.7470	3.30E-248	Formate--tetrahydrofolate ligase	Catalytic enzyme
HSP90	PF00183.14	9698 | m.11268	1.50E-244	Hsp90 protein	Protein folding and degredation
PGI	PF00342.15	10843 | m.12408	1.80E-243	Phosphoglucose isomerase	Fructose 6-phosphate formation
Menin	PF05053.9	4792 | m.6050	1.10E-241	Menin	Tumor suppressor gene
Ion_trans	PF00520.27	75 | m.156	7.30E-241	Ion transport protein	Ion transport
KAP	PF05804.8	6238 | m.7653	6.00E-239	Kinesin-associated protein (KAP)	Protein folding
Strabismus	PF06638.7	440 | m.716	1.80E-238	Strabismus protein	Unknown
PROCN	PF08083.7	288 | m.499	6.00E-234	PROCN (NUC071) domain	Pre-mRNA splicing
GDI	PF00996.14	11571 | m.13119	3.00E-232	GDP dissociation inhibitor	Regulate vesicular membrane trafficking
HSNSD	PF12062.4	116 | m.223	1.10E-230	heparan sulfate-N-deacetylase	Deacetylaition
eIF-3c_N	PF05470.8	6319 | m.7743	1.30E-229	Eukaryotic translation initiation factor 3 subunit 8 N-terminus	Early step of translation initiation
SNF	PF00209.14	3696 | m.4811	1.30E-229	Sodium:neurotransmitter symporter family	Release, re-uptake and recycling of neurotransmitters
RPE65	PF03055.11	17309 | m.18412	2.60E-102	Retinal pigment epithelial membrane protein	Cleavage of carotenoids
Pro-MCH	PF05824.8	13294 | m.14773	0.43	Pro-melanin-concentrating hormone (Pro-MCH)	Stimulates hunger and may additionally regulate energy homeostasis, reproductive function, and sleep

3.4 Conclusions:

Phenotypic polymorphisms and striking visual phenotypes have fascinated scientists for many years. The breadth of evolutionary causes for the maintenance of these phenotypes are as numerous as the species that display them. One organism, *Harmonia axyridis*, provides a unique opportunity to explore the genetic basis behind the maintenance of an easy to quantify variation - elytral spot number. This study provides and analyzes the first reference transcriptome completed for *H. axyridis*.

This study indicates that most gene expression profiles are shared across life stages of *H. axyridis.* While the majority of proteins identified in the assembled transcriptomes were structural in function, analyses of protein families using the Pfam database indicated the presence of pigment proteins in the adult, larva and optimized Vsearch merged transcriptomes. In particular, RPE65, which functions in the cleavage of carotenoids, was found in all assemblies. In *H. axyridis*, increased carotenoid pigmentation has been linked to increased alkaloid amounts (Britton et al., 2008). In addition, the elytral coloration of the seven spot ladybug, *Coccinella septempunctata*, is a result of several carotenoids (Britton et al., 2008). While larva are mostly black (**Figure 3.1**), we posit that the orange sections on the lower back could be due to carotenoid production. In addition, the melanin produce protein family found in the larval stage could be responsible for the black coloration.

The source of the orange coloration, as well as the function of the melanin family, should be explored in future studies, which are two of numerous research questions available in this new study species. Moreover, this study provides a necessary foundation for the continued study of the genetic link between genes and the maintenance of variation in *H. axyridis.* Chapter 4 discusses the development and analysis of a draft genome for *H. axyridis*, which is the next logical future direction to explore the genetic underpinnings of elytral spot and color polymorphisms.

CHAPTER IV- DRAFT GENOME CONSTRUCTION AND ANALYSIS

To further explore the evolution of phenotypic variation in *Harmonia axyridis*, a draft genome is essential, and this chapter describes data to pursue this future research direction. The following steps below have been undertaken to build a draft genome, and the pipeline for the data analyses is shown in **Figure 4.1**.

1. Sequencing with multiple platforms: Illumina and Nanopore
2. Initial genome assembly with multiple assembly algorithms: ABySS, SPAdes, WGS, L_RNA_Scaffolder and BESST_RNA
3. Evaluation of assembly completeness: BUSCO

The chapter is outlined as follows:

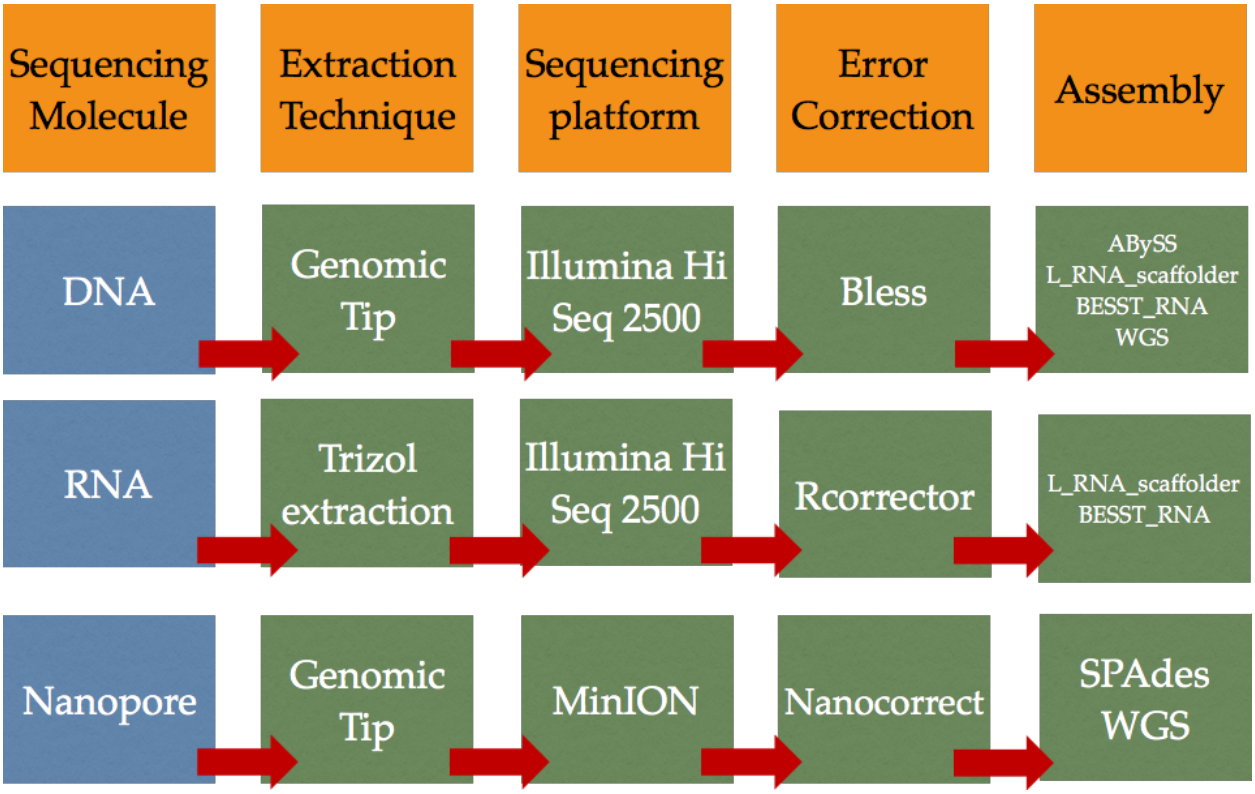

Figure 4.1: Sequencing and assembly pipeline

4.1 DNA extraction, sequencing and preliminary qualification (Illumina)

DNA was extracted from a single orange adult *Harmonia axyridis* using the Genomic Tip Extraction protocol (QUIAGEN). Extracted DNA was sent to the Hubbard Genome center at the University of New Hampshire for Illumina sequencing using a HiSeq 2500 sequencer. This produced 80 million-paired end, 150 base pair reads with an estimated coverage of 40X. The coverage was estimated using the equation: C = (read length * number of reads)/ genome size.

PreQC, which is part of the SGA assembler package, was used for an initial read analysis (Simpson J. T., 2014). PreQC provides the user with a variety of read statistics that help identify potential assembly pitfalls (https://goo.gl/R3HAzR).

The estimated genome size provided by PreQC was 827.9 MegaBases (**Figure 4.2A**). This genome size estimate was used to compare different genome assembly algorithm outputs. This is a potential assembly quality evaluator because significant differences between the assembly size and the actual genome size prediction made by PreQC would indicate an incomplete genome.

In addition to estimated genome size, variant branches provide insight into the heterozygosity of the genome (**Figure 4.2B**). The higher the

heterozygosity, the more difficult the genome is to assemble because such variants increase de Bruijn graph complexity, which is difficult for the assembler to resolve (Kajitani et al, 2014). From PreQC analysis, it was determined that *Harmonia axyridis* has a slightly heterozygous genome, because it is 10 fold less heterozygous than the oyster genome, which is highly heterozygous (Simpson, 2013).

There are a large number of repeat units within the genome (**Figure 4.2C**). These repeat units led to challenges during the genome assembly process. The GC content was around 30% (**Figure 4.2D**), which, based on examination of CG content across multiple animals, does not vary significantly from expectations (Amit et al., 2012).

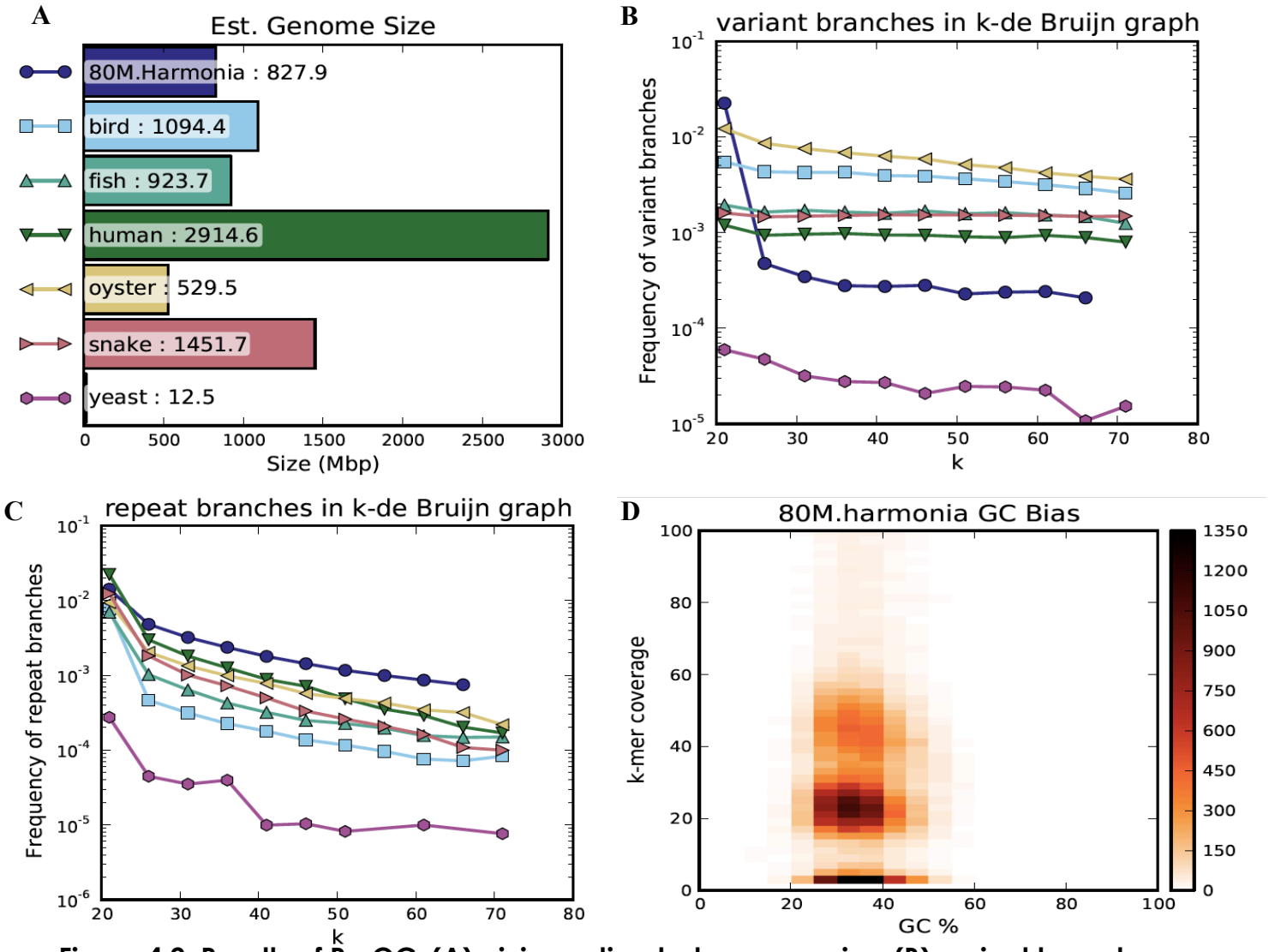

Figure 4.2: Results of PreQC: (A) giving estimated genome size, (B) variant branches, (C) repeat branches and (D) GC bias. The *Harmonia* genome is in dark blue.

4.2 RNA extraction and sequencing

RNA from four unfed adults (**Figure 4.3**) was extracted following the TRIzol extraction protocol (Invitrogen, Carlsbad USA). Unfed individuals were used in an attempt to reduce *A. pisum* contamination. The quantity and quality of extracted RNA was analyzed using a Qubit (Life Technology, Carlsbad USA) as well as a Tapestation 2200 (Agilent technologies, Palo Alto USA) prior to library construction. Following verification, extracted RNA was constructed into cDNA libraries for all samples following the TruSeq stranded RNA prep kit protocol (Illumina, San Diego USA). To allow multiple samples to be sequenced in a single lane, a unique adapter was ligated to each sample. These samples were then pooled in equimolar quantities. The cDNA libraries were sent to the New York Genome Center (New York, USA) for sequencing on a single lane (125bp paired end) of the HiSeq 2500 sequencer.

Figure 4.3: The four morphs from which RNA was extracted

4.3 DNA extraction and sequencing (Nanopore)

Separate Nanopore libraries were constructed with using DNA extracted from two adult *Harmonia axyridis* using the Genomic tip extraction protocol. One of the specimens was a newly emerged (from the pupal casing) 16-spot female. The second specimen was a yellow-orange 19 spot male. Preceding library construction, a Qubit (Life Technology, Carlsbad USA) was used to analyze DNA concentration. Libraries for the MinION sequencer were constructed using the Nanopore Sequencing Kit protocol SQK-MAP005. The protocol was followed as written, with the exception of DNA shearing, which was not performed.

Figure 4.4: *H. axyridis* used for MinION DNA extraction and sequencing. The top image is the first individual used, the bottom image is the second specimen used

4.4 Read processing and assembly

Read processing and de novo assembly-DNA (Illumina)

Prior to assembly, Illumina reads were error corrected using BLESS version 0.17 (Heo et al., 2014) (https://goo.gl/7SCYpz). Following Bless, the forward and reverse reads were interleaved into one file for downstream analysis (https://goo.gl/2dnB8O).

To hasten the assembly process, the interleaved files were split into 22 separate files (https://goo.gl/2dnB8O). ABySS (Simpson et al., 2009) was used to assemble to error corrected reads. To ensure the optimal assembly possible, kmer length was set to 91, 101, 111, and 121 using a shell script (https://goo.gl/Y6Gu0q).

The NG50 value of the multiple ABySS runs indicated that a kmer length of 111 was optimal for the *H. axyridis* ABySS assembly; therefore and this assembly was used in downstream analysis.

Read processing – cDNA libraries (Illumina)

The four separate cDNA libraries were error corrected using Rcorrector (Version 3 - Song and Florea, 2015) following suggestions outlined in MacManes, (2015; https://goo.gl/pFVTzr). The default settings were used, with the exception of using 10 threads to decrease processing time. These error-corrected reads were used in downstream analysis.

Read processing and de novo assembly- DNA (Nanopore)

Two separate Nanopore libraries were run on a MinION Nanopore sequencer. Poretool stats (Loman and Quinlan, 2014) was used to analyze raw read data (https://goo.gl/X6xTQu) and Poretools fastq was used to convert the files to fastq form. To filter out short reads, the min-length was set to 1,000 (https://goo.gl/Hhzhzi). Following analysis, the two run outputs were concatenated (https://goo.gl/h5wN0L). Nanocorrect (Loman, Quick and Simpson, 2015) was used to correct the combined Nanopore file for SPAdes assembly (https://goo.gl/kxbAZa). Initially, SPAdes (version 3.5) (Bankevich et al., 2012) was used to assemble the corrected Nanopore reads together with the Illumina data (https://goo.gl/qwPHx1).

Assembly using Illumina and Nanopore reads

SPAdes led to a short and incomplete genome. In light of this, WGS (Myers et al., 2000), an alternative assembly algorithm, was run. To prepare the reads for WGS, Nanocorrect (Loman, Quick and Simpson, 2015) was run on both Nanopore outputs (https://goo.gl/ncxlZK). These Nanocorrected fastq reads were converted to fasta reads (https://goo.gl/IkE9PZ). In order to prepare reads for WGS, the fasta files were then converted into .frg files (https://goo.gl/mA43mN)

Using the Nanopore data, MaSuRCA was used to create super reads of the Illumina data (https://goo.gl/fOTmvi). Finally, WGS (Myers et al., 2000) was

used to assemble the super-reads using both the first and second run error-corrected Nanopore reads. Within the specification file, the kmer length was set to 22(https://goo.gl/jmfHgE).

Assembly using Illumina, DNA and RNA reads

To improve assembly length and completeness, L_RNA_scaffolder (Xu et al., 2013) and BESST_RNA (https://github.com/ksahlin/BESST_RNA) were run. Both of these assembly algorithms use transcripts to scaffold genomes. BESST_RNA was run iteratively using a python script, meaning that the output of one run was used as the input of the next.

Prior to the assembly process, Blat (version 35X1, Kent, 2002) was run. Blat is an alignment tool that uses the genome as an index. Blast was run using the ABySS assembly and the final Vsearch merged transcriptome. This output and the ABySS assembled genome was then piped into L_RNA_scaffolder (https://goo.gl/Xcwp0B , https://goo.gl/1aADLW).

The four Rcorrected Illumina transcript files, extracted from four different phenotypic morphs (**Figure 4.2**), were aligned to the L_RNA_scaffolder assembly using BWA (https://goo.gl/oblvO0). The four alignment files were merged, and this merged file, along with the ABySS assembled genome, was used as an initial input into BESST_RNA (https://goo.gl/pDyqPG). The Illumina transcripts were aligned back to the BESST_RNA assembled genome and run through BESST_RNA again.

4.5 Assembly evaluation

The NG50 value and size of the output file were used as an initial assembly evaluation. In addition, BUSCO (version 1.1b1, Myers et al., 2000) was used for additional analysis (**Table 4.4**). This program analyzes completeness by comparing genes present in the assembly to a set of known genes present across 90% of species within taxonomically similar groups. The database used in assembly evaluations was Arthropoda (https://goo.gl/iMFlt7, https://goo.gl/WyLOLZ, https://goo.gl/TVmHwD, https://goo.gl/x3ll9J, https://goo.gl/iMFlt7).

Table 4.1: Completeness evaluation of the different assemblies and sequencing technology.

Technology	Assembler	N50	NG50	BUSCO Arthropoda	Size	Number of scaffolds
Illumina - DNA	ABySS	4700	5426	58%	846M	5.41E+05
Nanopore	SPAdes	2238	1138	34%	660M	4.69E+05
Illumina-DNA and Nanopore	WGS	4175	1051	10%	329M	8.69E+04
Illumina - DNA and RNA	L_RNA_scaffolder	4163	4052	57%	798M	5.12E+05
Illumina-DNA and RNA	BESST_RNA	4401	4211	35%	812M	5.01E+05

4.6 Chapter Summary:

This data chapter describes the assembly and evaluation of a draft genome for *Harmonia axyridis.* This will serve as a tool for future studies to explore the evolution of phenotypic variation in *H. axyridis*. To optimize the assembly process, multiple types of sequencing were used. Illumina sequencing was performed on both DNA and cDNA libraries, and Nanopore technology was used to sequence long strands of DNA.

The two most complete genomic assemblies were achieved using Illumina sequencing of DNA and cDNA. The assembly utilizing only DNA and the assembler ABySS had a completeness score of 59%. The assembly that used both Illumina RNA and DNA sequencing and the assembler L_RNA_scaffolder also had a completeness score of 59%. Since the length of the scaffolds created with ABySS was larger than the scaffolds found using L_RNA_scaffolder, the AbySS assembly should be the assembly used in downstream analysis. Surprisingly, the least complete assembly was the WGS assembly that utilized the Nanopore and Illumina DNA reads. This could be a result of the Nanopore data not producing sufficient data to aid the assembler during scaffolding. Surprisingly, the least complete assembly was the WGS assembly that utilized the Nanopore and Illumina DNA reads. This could be an artifact of the Nanopore data not producing sufficient data to aid the assembler during scaffolding.

For the two best performing assemblies, the NG50 and completeness assessment is lower than desired. However, the inclusion of additional Illumina and Nanopore sequencing in future work would allow the assembler to more accurately predict and bridge unknown gaps. This additional sequencing data would also help the assembler bridge the many repeat units present in the *H. axyridis* genome (**Figure 4.2C**).

4.7 Conclusions:

This dissertation provides a foundation for utilizing *Harmonia axyridis* as a model species in the study of the evolution of phenotypic variation. This was accomplished by quantifying polymorphisms in a local population, establishing a laboratory population for developmental observations, assembling a transcriptome of the adult and larval stages, and developing a draft genome for future work.

LIST OF REFERENCES

CHAPTER I - INTRODUCTION:

Alam N., Choi I. S., Song K-S., Hong J., Lee C. O., Jung J. H. (2002). A New Alkaloid from Two Coccinellid Beetles *Harmonia axyridis* and *Aiolocaria hexaspilota*.*Bull. Korean Chem. Soc.* 23:497-499.

Awad M., Laugier G. J. M., Loiseau A., Nedvéd. (2015). Unbalanced polyandry in wild-caught ladybirds Harmonia axyridis (Coleoptera: Coccinellidae). *Appl. Entomol. Zool* DOI 10.1007/s13355-015-0348-5.

Bind R. B. (2007). Reproductive behavior of a generalist aphidophagous ladybird beetle *Cheilomenes sexmaculata* (Coleoptera: Coccinellidae). *Internat. J. Trop. Insect Sci.* 27: 78-84.

Brakefield P. M., de Jong P. W. (2011). A steep cline in ladybird melanism has decayed over 25 years: a genetic response to climate change? Heredity 107: 574-578.

Brown P.M.J., Adriaens T., Bathon H., Cuppen J., Goldarazena A., Hagg T., Kenis M., Klausnitzer B. E.M., Kovar I., Loomans A. J. M., Majerus M. E. N., Nedved O., Pedersen J., Rabitsch W., Roy H. E., Ternois V., Zakharov I. A., Roy D. B. (2007). *Harmonia axyridis* in Europe: spread and distribution of a non-native coccinellidea. BioControl 53: 5-22.

Cadwell J. P., (1996). The Evolution of myrmecophagy and its correlates in poison frogs (Family Dendrobatidae). *Journal of zoology* 240: 75-101.

Dobzhansky T. (1993). Geographical variation in ladybeetles. The American Naturalist 67: 97-126

Eizirik E., Yuhki N., Johnson W.E., Menotti-Raymond M., Hannah S. S., O'Brien S. J. (2003). Molecular Genetics and Evolution of Melanism in the Cat Family. *Curr. Biol* 13: 448-453.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Flanagan N.S., Tobler A., Davison A., Pybus O.G., Kapan D. D., Planas S., Linares M., Heckel D., McMillan W. O. (2004). Historical demography of Müllerian mimicry in the neotropical *Heliconius* butterflies. *PNAS* 101: 9704-9709.

Forsman A., Karlsson M., Wennersten L., Johansson J., Karpenstam E. (2011). Rapid evolution of fire melanism in replicated populations of pygmy grasshoppers. *Evolution* 65:2530-2540.

Gittleman J. L., Harvey P. H., Greenwood P. J. (1980). The evolution of conspicuous coloration: some experiments in bad taste. *Anim. Behav* 28**:** 897-899.

Gordon R.D. (1985). The Coleoptera (Coccinellidae) of America north of Mexico. J New York Entomol Soc. 93:1–912

Hales D.F., Tomiuk J., Woehrmann K., Sunnucks P. (1997). Evolutionary and genetics aspects of aphid biology: A review. Eur. J. Entomol, 94 (1): 1-55.

Hodek I., van Emden H. F., Honêk A. (2012). Ecology and Behavior of the Ladybird Beetles (Coccinellidae). Oxford: Wiley-Blackwell.

Hoekstra H. E., Hirschmann R. J., Bundey R. A., Insel P. A., Crossland J. P. (2006). A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern. *Science* 313: 101-104.

Jones R.T., Salazar P. A., ffrench-Constant R. H., Jiggins C.D., Joron M. (2012). Evolution of a mimicry supergene from a multilocus architecture. *Proc. R. Soc. B* 279: doi: 10.1098/rspb.2011.0882.

Joron M., Jiggins C. D., Papanicolaou A., McMillan W. O. (2006). *Heliconius* wing patterns: an evo-devo model for understanding phenotypic diversity. *Heredity* 97: 157-167.

Karpestam E., Merilaita S., Forsman A. (2014). Natural levels of colour polymorphism reduce performance of visual predators searching for camouflaged prey. *Biol J Linnean Soc.* doi: 10.1111/bij.12276.

Koch R.L. (2003). The multicolored Asian lady beetle, Harmonia axyridis: A review of its biology, uses in biological control, and non-target impacts. *Journal of Insect Science* 3: 32-48.

Kronforst M. R., Young L. G., Kapan D. D., McNeely C., O'Neill R. J., Gilbert L. E. (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. *PNAS* 103: 6575-6580.

LaMana M.L., Miller J.C. (1996). Field observations on *Harmonia axyridis* Pallas (Coleoptera: Coccinellidae) in Oregon. Biological Control 6: 232-237.

Lande R. (1981). Models of speciation by sexual selection on polygenic traits. *Evolution* 78: 3721-3725.

Lindström L., Alatalo R. V., Mappes J. (1997). Imperfect Batesian mimicry – the effects of the frequency and the distastefulness of the model. *Proc. R. Sco. Lond. B.* 264: 149-153.

Lombaert E., Malausa T., Davred R., Estoup A. (2008). Phenotypic variation in invasive and biocontrol populations of the harlequin ladybird, *Harmonia axyridis*. *BioControl* 53: 89-102.

Loyau A., Jalme M. S., Cagniant C., Sorci G. (2005). Multiple sexual advertisements honestly reflect health status in peacocks (*Pavo cristatus*). *Behav Ecol Sociobiol* 58: 552-557.

Müller F. (1879). *Ituna* and *Thyridia*: a remarkable case of mimicry in butterflies (R. Meldola, translation.). *Proc. Entomol.Soc. Lond.* 1878:20-29.

Omkar, Singh S. K. (2005). Mating behavior of an aphidophagous ladybird beetle, *Propylae dissecta* (Mulsant). *Insect Sci.* 12: 37-44.

Osawa N., Nishida T. (1992). Seasonal variation in elytral colour polymorphism in *Harmonia axyridis* (the ladybird beetle): the role of non-random mating. Heredity 69: 297-307.

Perry J. C., Roitberg B. D. (2005). Ladybird mothers mitigate offspring starvation risk by laying trophic eggs. *Behav. Ecol. Sociobiol.* 58: 578-586.

Pfennig D. W., Harcombe W. R., Pfennig K. S. (2001). Frequency-dependent Batesian mimicry. *Nature* 410: 323.

Santos J. C., Coloma L. A., Cannatella D. C. (2003). Multiple, recurring origins of aposematism and diet specialization in poison frogs. PNAS: doi: 10.1073/pnas.2133521100

Servedio M. R. (2000). The effects of predatory learning, forgetting, and recognition errors in the evolution of warning coloration. *Evolution* 54: 751-763.

Sword G.A (2002). A role for phenotypic plasticity in the evolution of aposematistm. *Proceedings of the Roryal Soceity of London* B 269:1639-1644.

Tsurui K., Honma A., Nishida T. (2010). Camouflage Effects of Various Colour-Marking Morphs against Different Microhabitat Backgrounds in Polymorphic Pygmy Grasshoper *Tetrix japonica*. *PLOSONE*: DOI: 10.1371/journal.pone.0011446

Vignieri S. N., Larson J .G., Hoekstra H. E. (2010). The selective advantage of crypsis in mice. DOI: 10.1111/j.1558-5646.2010.00976.x

Ware R. L.,Yguel B., Majerus M. E. N. (2008). Effects of larval diet on female reproductive output of the European coccinellid *Adalia bipunctata* and

the invasive species *Harmonia axyridis* (Coleoptera: Coccinellidae). *Eur. J. Entomol.* 105: 437-443.

CHAPTER II – ESTABLISMENT OF A MATING COLONY OF *HARMONIA AXYRIDIS* IN A LABORATORY SETTING

Brown M. W., Miller S. S. (1998). Coccinellidae (Coleoptera) in apple orchards of eastern West Virginia and the impact of invasion by Harmonia axyridis. *Entomological News* 109: 143-151.

Brown P. M. J., Thomas C. E., Lombaert E., Jeffries D. L., Estoup A., Handley L-J. L. (2011). The global spread of *Harmonia axyridis* (Coleoptera: Coccinellidae): distribution, dispersal and routes of invasion. *BioControl* 56: 623-641.

Cailaud M.C., Boutin M., Braendle C., Simon J. C. (2002). A sex-linked locus controls wing polymorphism in males of pea aphid, *Acyrthosiphon pisum* (Harris). *Heredity* 89: 346-352.

Denlinger D. L. (2002). Regulation of diapause. *Annu. Rev. Entomol.* 47: 93-122.

Elliot N C., Kieckhefer R. W., Kauffman W. C. (1991). Estimating adult coccinellid populations in wheat fields by removal, sweepnet, and visual count sampling. *Can. Entomol.* 123:13-22.

Hardie J., Lees A. D. (1985). Endocrine control of polymorphism and polyphenism. *Comprehensive Insect Physiology, Biochemistry and Pharmacology*: 8: 441-449.

Hodek I., Cerkasov J. (1961). Prevention and artificial induction of the imaginal diapause in *Coccinella 7-punctata*. *Ent. Exp & appl* 4: 179-190.

Hodek I., van Emden H. F., Honêk A. (2012). Ecology and Behavior of the Ladybird Beetles (Coccinellidae). Oxford: Wiley-Blackwell.

Kidd K. A., Nalepa C. A., Day E. R., Waldvogel M. G. (1995). Distribution of *Harmonia axyridis* (Palla) (Coleoptera: Coccinellidae) in North Carolina and Virginia. *Proc. Entomol. Soc. Wash.* 97: 729-731.

Majerus M., Strawson V., Roy H. (2006). The potential impact s of the arrival of the harlequin ladybird, *Harmonia axyridis* (Pallas) (Coleoptera: Coccinellidae) in Britain. *Ecological Entomology* 31: 207-215.

Nakazawa T., Satinover S. M., Naccara L., Goddard L., Dragulev B. P., Peters E., Platts-Mills T. A.E. (2007). Asian ladybugs (*Harmonia axyridis*): A new seasonal indoor allergen. *Journal of Allergy and Clinical Immunology* 119:421-427.

Raak-Van den Berg C. L., De Jong P. W., Hemerik L., Van Lenteren J. C. (2013). Diapause and post-diapause quiescence demonstrated in overwintering *Harmonia axyridis* (Colopertera: Coccinellidae) in northwestern Europe. *Eur. J. Entomol* 110: 585-592.

Raak-van den Berg C. L., Hemerik L., de Jong P. W., van Lenteren J. C (2012). Mode of overwintering of invasive *Harmonia axyridis* in the Netherlands. *BioControl* 57:71-84.

Ueno H. (1996). Estimation of Multiple Insemination in a Natural Population of *Harmonia axyridis* (Coleptera:Coccinellidae). *Appl. Entomol. Zool* 21: 621-623.

Via S. (1999). Reproductive Isolation between Sympatric Races of Pea Aphids. I. Gene Flow Restriction and Habitat Choice. *Evolution* 53: 1446-1457.

CHAPTER III- CHARACTERIZATION OF THE TRANSCIPTOME OF THE MULTICOLORED ASIAN LADY BEETLE, *HARMONIA AXYRIDIS*

Ashburner M., Ball C.A., Blake J. A., Botstein D., Butler H., Cherry M., Davis A. P., Dolinski K., Dwight S. S., Epping J.T., Harris M. A., Hill D. P., Tarver L. I., Kasarskis A., Lewis S., Matese J. C., Richardson J. E., Ringwald M., Rubin G. M., Sherlock G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics* 25: 25-29.

Bezzerides A. L., McGraw K. J., Parker R. S., Husseini J. (2007). Elytral color as a signal of chemical defense in the Asian ladybird beetle *Harmonia axyridis*. *Behavior Ecology and Sociobiology* 61: 1401-1408.

Bolger A. M., Lohse M., Usadel., B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Bray, N., Pimentel, H., Melsted, P. & Pachter, L. (2015) Near-optimal RNA-Seq quantification. arXiv preprint arXiv:1505.02710.

Britton G., Liassen-Jesen, S., Pfander H.(2008). Carotenoids, Vol. 4: Natural Functions. *Springer.* DIO: 0.1007/978-3-7643-7499-0

Cadwell J. P., (1996). The Evolution of myrmecophagy and its correlates in poison frogs (Family Dendrobatidae). *Journal of zoology* 240: 75-101.

Flanagan N.S., Tobler A., Davison A., Pybus O.G., Kapan D. D., Planas S., Linares M., Heckel D., McMillan W. O. (2004). Historical demography of Müllerian mimicry in the neotropical *Heliconius* butterflies. *PNAS* 101: 9704-9709.

Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles D., Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Wasterman R., William T., Dewey C. N., Henschel R., LeDuc R. D., Friedman N., Regev A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using Trinity platform for reference generation an analysis. *Nature Protocols* 8: 1494-1512.

Heo Y., Wu X., Chen D., Ma J., H W. (2013). BLESS: Bloom filtered-based error correction solution for high-throughput sequencing reads. *Bioinformatics* doi: 10.1093/bioinformatics/btu030.

Jones R.T., Salazar P. A., ffrench-Constant R. H., Jiggins C.D., Joron M. (2012). Evolution of a mimicry supergene from a multilocus architecture. *Proc. R. Soc. B* 279: doi: 10.1098/rspb.2011.0882.

Joron M., Jiggins C. D., Papanicolaou A., McMillan W. O. (2006). *Heliconius* wing patterns: an evo-devo model for understanding phenotypic diversity. *Heredity* 97: 157-167.

Karpestam E., Merilaita S., Forsman A. (2014). Natural levels of colour polymorphism reduce performance of visual predators searching for camouflaged prey. *Biol J Linnean Soc.* doi: 10.1111/bij.12276.

Kronforst M. R., Young L. G., Kapan D. D., McNeely C., O'Neill R. J., Gilbert L. E. (2006). Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. *PNAS* 103: 6575-6580.

MacManes M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Front Genet* 5: 13. doi: 10.3389/fgene.2014.00013.

Osawa N., Nishida T., (1992). Seasonal variation in elytral colour polymophims in *Harmonia axyridis* (the ladybird beetle): the role of non-random mating. *Heredity* 69: 297-307.

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 dio: 10.1093/bioinformatics/btv351.

Smith-Unna R. D., Boursnell C., Patro R., Hibberd J. M., Kelly S. (2015). TransRate: reference free quality assessment of *de-novo* transcriptome assemblies. *bioRxiv*: dio http://dx.doi.org/10.1101/021626.

CHAPTER IV – DRAFT GENOME CONSTRUCTION AND ANALYSIS

. Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome research*, 19(6), 1117-1123.

Amit M., Donyo M., Hollander D., Goren A., Kim E., Gelfman S., Lev-Maor G., Burstein D., Schwartx S., Postolsky B., Pupko T., Ast G. (2012). Differential GC Content between Exons and Introns Establishes Distinct Strategies of Splice-Site Recognition. *Cell Reports.* 1(5): 543-556.

Bankevich A., Nurk S., Antipov D., Gurevich A. A., Dvorkin M., Kulikov A. S., Lesin V. M., Nikolenko S. I., Pham S., Prjibelski A. D., Pyshkin A. V., Sirotkin A. V., Vyahhi N., Tesler G., Alekseyev M. A., Pevzner P.A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput Biol.* 19(5): 455-477.

Heo Y., Wu X., Chen D., Ma J., H W. (2013). BLESS: Bloom filtered-based error correction solution for high-throughput sequencing reads. *Bioinformatics* doi: 10.1093/bioinformatics/btu030.

Kent W. J. (2002). BLAT- the BLAST-like alignment tool. *Genome Res.* 12(4): 656-664.

Loman N. J., Quick J., Simpson J.T. (2015). A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nature methods*. 12: 733-735.

Loman N. J., Quinlan A. R. (2014). Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics*. 30(23): 3399-3401.

MacManes, M.D. Optimizing error correction of RNAseq reads. bioRxiv doi: http://dx.doi.org/10.1101/020123.

Myers E.W., Sutton G.G., Delcher A. L., Dew I. M., Fasulo D. P., Flanigan M. J., Kravitz S. A., Mobarry C. M., Reinert K. H., Remington K. A., Anson E. L., Bolanos R. A., Chou H. H., Jordan C. M., Halpern A. L., Lonardi S., Beasley E. M., Brandon R. C., Chen L., Dunn P. J., Lai Z., Liang Y., Nusskern D. R., Zhan M., Zhang Q., Zeheng X., Rubin G.M., Adams M. D., Venter J. C. (2000). A whole-genome assembly of Drosophila. *Science*. 287(5461): 2196-2204.

Simpson J. T. (2014). Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* 30(9): 1228-1235.

Song, L., Florea, L. (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. GigaScence. 4:48.

Xue E., Li J-T., Zhu Y-P., Hou G-Y., Kong X-F., Kuang Y-Y., Sun X-W. (2013). L_RNA_scaffolder: scaffolding genomes with transcripts. *Bioinformatics* 14:604.