Characterizing the Adult and Larval Transcriptome 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

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*.

32

Introduction:

34 The evolution and maintenance of phenotypic polymorphism and striking visual phenotypes have
35 fascinated scientists for many years (Darwin, 1859; Endler, 1986; Fisher, 1930; Gray and McKinnon,
36 2006; Joron et al., 2006). In general, insects have become increasingly popular as study organisms to
37 examine phenotypic variation (Jennings, 2011; Joron et al., 2006). One such insect displaying extensive
38 elytra and spot variation that has yet to be extensively studied is the Asian Multicolored Ladybeetle,
39 *Harmonia axyridis*.

40

41 The mechanisms responsible for the evolution of these phenotypes are as widespread as the species
42 that display them. Aposematism, crypsis, and mimicry may play a role in the evolution of phenotypic
43 variation in the animal kingdom. Members of family *Dendrobatidae,* poison dart frogs, are
44 aposematically colored (Cadwell, 1996), while *Tetrix subulata* grasshoppers maintain their phenotypic
45 polymorphism to aid in crypsis (Karpestam et al., 2014). A mimicry strategy is utilized by one particularly
46 well-characterized species that exhibits phenotypic polymorphism, the Neotropic butterfly system,
47 *Heliconius*. The color, pattern, and eyespot polymorphism seen in *Heliconius* is thought to have arose
48 as a result of Müllerian mimicry (Flanagan et al., 2004) and the supergenes underlying these traits have
49 been well characterized (Kronforst et al., 2006, Joron et al., 2006, Jones et al., 2012).

50

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

58

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

67

68   Sexual selection may play a role in color variation in *H. axyridis.* Osawa and Nishida (1992) remarked
69   that female *H. axyridis* might choose their mates based on melanin concentration. Their choice,
70   however, has been shown to vary based on season and temperature. Non-melanic (red, orange, or
71   yellow with any spot number) males have a higher frequency of mating in the spring-time, while melanic
72   (black) males have an increased frequency of mating in the summer. While this has been shown with
73   respect to elytral color, no such findings have occurred for spot number. Although these spot patterns
74   are believed to be related to predator avoidance, thermotolerance, or mate choice (Osawa and Nishida,
75   1992), the genetics underlying these patterns is currently unknown.

76

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

80

81   **Methods and Materials:**

82   ***Specimen capture, RNA extraction, library prep and sequencing***

83

84   One larval (Figure 1a) and one adult (Figure 1b) *H. axyridis* were captured on the University of New
85   Hampshire campus in Durham, New Hampshire (43.1339° N, 70.9264° W). The adult was orange with
86   18 spots. The insects were immediately placed in RNAlater and stored in a -80C freezer until RNA
87   extraction was performed. The RNA from both individuals was extracted following the TRIzol extraction
88   protocol (Invitrogen, Carlsbad USA). The entire insect was used for the RNA extraction protocol. The
89   quantity and quality of extracted RNA was analyzed using a Qubit (Life Technology, Carlsbad USA) as
90   well as a Tapestation 2200 (Agilent technologies, Palo Alto USA) prior to library construction. Following
91   verification, RNA libraries were constructed for both samples following the TruSeq stranded RNA prep
92   kit (Illumina, San Diego USA). To allow multiple samples to be run in one lane, a unique adapter was
93   added to each sample. These samples were then pooled in equimolar quantities. The libraries were
94   then sent to the New York Genome Center (New York, USA) for sequencing on a single lane (125bp
95   paired end) of the HiSeq 2500 sequencer.

96

97

98

Figure 1a: The larva used for transcriptome sequencing.

100

101

102

Figure 1b: The adult used for transcriptome sequencing.

104

105

Sequence Data Preprocessing and Assembly

107

The raw sequence reads corresponding to the two tissue types were error corrected using the software BLESS (Heo et al., 2014) version 0.17 (https://goo.gl/YHxlzI, https://goo.gl/vBh7Pg). The error-corrected sequence reads were adapter and quality trimmed following recommendations from MacManes (MacManes, 2014) and Mbandi (Mbandi, 2014). Specifically, adapter sequence contamination and low quality nucleotides (defined as Phred<2) were removed using the program

113   Trimmomatic version 0.32 (Bolger, 2014) called from within the Trinity assembler version 2.1.1 (Haas,

114   2013). Reads from each tissue were assembled together to created a joint assembly of adult and larva

115   transcripts using a Linux workstation with 64 cores and 1Tb RAM. We used flags to indicate the stranded

116   nature of sequencing reads and set the maximum allowable physical distance between read pairs to

117   999nt (https://goo.gl/ZYP08M).

118

119   The quality of the assembly was evaluated using transrate version 1.01 (Unna-Smith 2015;

120   (https://goo.gl/RpdQSU). Transrate generates quality statistics based on a process involving mapping

121   sequence reads back to the assembled transcripts. Transcripts supported by properly mapped reads of

122   a sufficient depth (amongst other things) are judged to be of high quality. In addition to generating quality

123   metrics, transrate produces an alternative assembly with poorly-supported transcripts removed. This

124   improved assembly was used for all downstream analyses and QC procedures. We then evaluated

125   transcriptome completeness via use of the software package BUSCO version 1.1b (Simão 2015).

126   BUSCO searches against a database of highly-conserved single-copy genes in Arthropoda

127   (https://goo.gl/bhTNdr). High quality, complete transcriptomes contain are hypothesized to contain the

128   vast majority of these conserved genes.

129

130   To remove assembly artifacts remaining after transrate optimization, we estimated transcript abundance

131   using 2 software packages - Salmon version 0.51 (Patro 2015; https://goo.gl/01UIF6) and Kallisto

132   version 0.42.4 (Bray, 2015; https://goo.gl/BsQMpr). Transcripts whose abundance exceeded 0.5 TPM

133   in either adult or larval datasets using either estimation method were retained. We evaluated

134   transcriptome completeness and quality, again, after TPM filtration, using BUSCO and transrate, to

135   ensure that our filtration processes did not significantly effect the biological content of the assembly.

136

137   ***Assembled Sequence Annotation***

138

139   The   filtered   assemblies   were   annotated   using   the   software   package   dammit

140   (https://github.com/camillescott/dammit; https://goo.gl/05MY5i). Dammit coordinates the annotation

141   process, which involves use of blast (Camacho, 2009), TransDecoder (version 2.0.1,

142   http://transdecoder.github.io/), and hmmer version 3.1b1 (Wheeler 2013). In addition to this, putative

143   secretory proteins were identified using the software signalP, version 4.1c (https://goo.gl/FaOQSj).

144

145   To identify patterns of gene expression unique to each life stage, we used the expression data as per

146   above. We identified transcripts expressed in one stage but not the other, and cases where expression

147   occurred in both life stages. The Uniprot ID was identified for each of these transcripts using a blastx

148   search (https://goo.gl/J9saMj), and these terms were used in the web interface Amigo (Carbon et al.,

149   2009) to identify Gene Ontology terms that were enriched in either adult or larva relative to the

150 background patterns of expression. The number of unique genes contained in the joint assembly was

151 calculated via a BLAST search of the complete gene sets of *Homo sapiens*, *Drosophila melanogaster*,

152 and *Tribolium casteneda*.

153

154 **Results and Discussion:**

155

156 ***Data Availability***

157

158 All read data are available via https://goo.gl/VNuHXC. Assemblies and data matrices are available at

159 https://goo.gl/D3xh65.

160

161 ***RNA extraction, Assembly and Evaluation***

162 RNA was extracted from whole bodies of both the adult and the larva stage of a single *Harmonia*

163 *axyridis*. The quality was verified using a Tapestation 2200 as well as a Qubit. The initial concentration

164 for the larva sample was 83.2 ng/uL, while the initial concentration for the adult sample was 74.7 ng/uL.

165 The number of strand-specific paired end reads contained in the adult and larva libraries were 58 million

166 and 67 million, respectively. The reads were 125 base pairs in length.

167

168 The raw Trinity assembly of the larval and adult reads resulted in a total of 171,117 contigs (82Mb)

169 exceeding 200nt in length. This assembly was evaluated using Transrate, producing an initial score of

170 0.10543, and and optimized score of 0.29729. This transrate optimized assembly (89,305 transcripts,

171 62Mb) was further filtered by removing transcripts whose expression was less than 0.5 TPM. After

172 filtration, 33,648 transcripts (40Mb) remained. To assess for the inadvertent loss of valid transcripts, we

173 ran BUSCO before and after this filtration procedure. The percent of Arthropoda BUSCO's missing from

174 the assembly rose slightly, from 18% to 21%. Transrate was run once again, and resulted in a final

175 assembly score of 0.29112. This score is indicative of a high-quality transcriptome appropriate for

176 further study. In an attempt at understanding how many distinct genes our transcriptome contained, we

177 conducted a blast search against *Homo sapiens*, *Drosophila melanogaster*, and *Tribolium casteneda.*

178 This search resulted in 7,246, 7,739, and 7,741 unique matches, which serve as estimates of the

179 number of unique genes expressed in these two life stages. The final assembly is available at

180 https://goo.gl/nWdBuv

181

182 ***Annotation***

183

184   The assembled transcripts were annotated using the software package dammit!, which provided
185   annotations for 23,304, or 69% of the transcripts (available here: https://goo.gl/gpGXLG). These
186   annotations included putative protein and nucleotide matches, 5- and 3-prime UTRs, as well as start
187   and stop codons. In addition to this, analysis with Transdecoder yielded 14,518 putative protein
188   sequences (available here: https://goo.gl/qVLWwD), which were annotated by 4,139 distinct Pfam
189   protein families, while 176 transcripts were determined to be non coding (ncRNA) based on significant
190   matches to the Rfam database (available here: https://goo.gl/x1n7jC). Lastly, 2,925 proteins (7.8% of
191   total) were determined to to be secretory in nature by the software package signalP (available here:
192   https://goo.gl/z0ra1g).

193

194   Annotation of the sequence dataset resulted in the identification of host of transcripts that may be of
195   interest to other researchers including: 43 heat-shock and 8 cold-shock transcripts, 87 homoebox-
196   domain containing transcripts, 122 7-transmembrane-containing (18 GPCR's) transcripts, 13 solute
197   carriers, 143 ABC-transport-containing transcripts, and 21 OD-S (pheromone-binding) transcripts.

198

199   A complement of immune-related genes were discovered as well. These include a single member of
200   the Attacins and Coleoptericins, two TLR-like genes, seven Group 1, and 34 Group 2 C-type lectin
201   Receptors (CLRs). Two CARD-containing Cytoplasmic pattern recognition receptor (CRR) genes were
202   discovered, as were 3 MAP kinase containing transcripts. Finally, 119 RIG-I-like receptors (RLR) were
203   found.

204

205   Gene expression was estimated for each transcript for both adult and larva (Figure 2, available here:
206   https://goo.gl/wM3TV7). For all transcripts expressed in the adult, the mean TMP=29.7 (max= 80,016.9,
207   SD= 500.8) while the mean larval TPM=29.71 (max= 166,264, SD=941). When analyzing transcripts
208   found uniquely in these two tissues, mean adult TMP=9.9 (max=3,037, SD=86) and for larva TPM=8.6
209   (max=687, SD=41).

1922 transcripts expressed in adult only

30630 transcripts expressed in both adult and larva

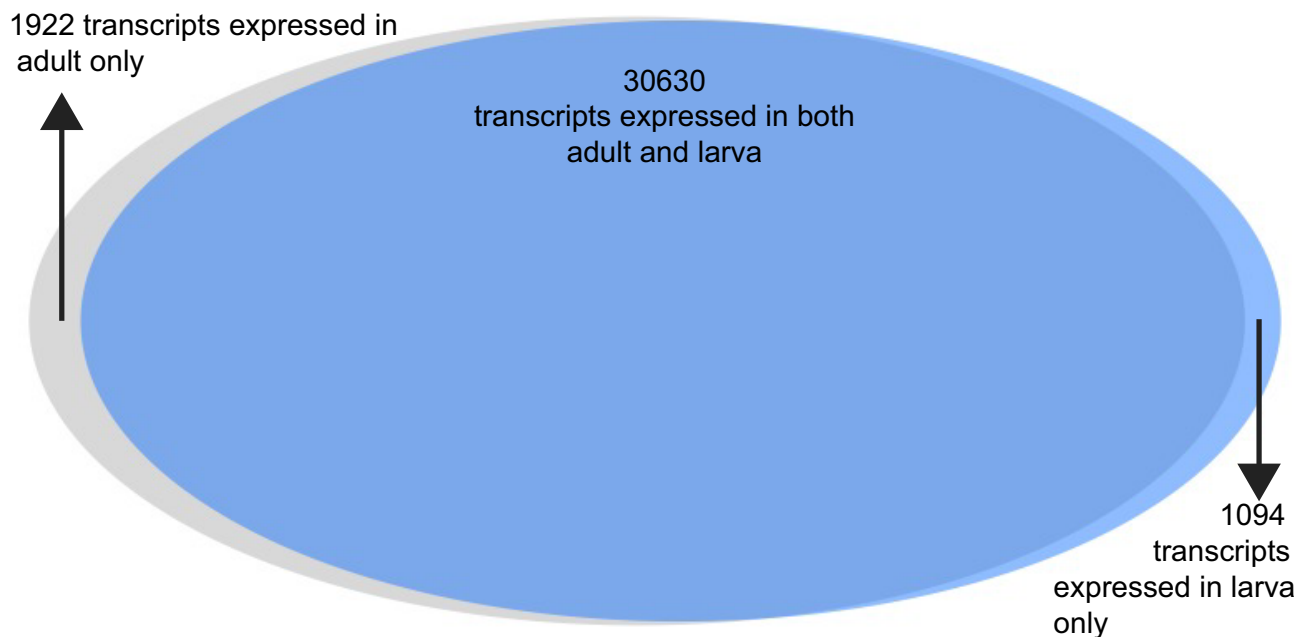1094 transcripts expressed in larva only

210

211

212 **Figure 2: The Venn diagram representing the number of transcripts expressed in both adult and**
213 **larva, as well as those expressed uniquely in one or the other.**

214

215 Analysis of the differences between adult and larval life stages were carried out as well. Because these
216 life stages were only sequenced with a single individual each, they should be interpreted with some
217 caution. The vast majority of transcripts were observed in both life stages (n=30,630, 91%), with a small
218 number being expressed uniquely in larva (n=1,094) and adult (n=1,922). Of these transcripts
219 expressed uniquely in either larva or adult, 6.1% and 4.6%, respectively, were found to be secretory in
220 nature.

221

222 **Conclusions:**

223 Phenotypic polymorphisms and striking visual phenotypes have fascinated scientists for many years.
224 The breadth of evolutionary causes for the maintenance of these phenotypes are as numerous as the
225 species that display them. One organism, *Harmonia axyridis*, provides a unique opportunity to explore
226 the genetic basis behind the maintenance of an easy to quantify variation - elytral spot number. While
227 understanding these genomic mechanisms is beyond the scope of this paper, we do provide a reference
228 transcriptome for *H. axyridis*, a foundational resource for this work.

229

230 This study indicates that most gene expression profiles are shared across life stages of *H. axyridis.*
231 While the majority of proteins identified in the assembled transcriptome were structural in function,
232 analyses of protein families using the Pfam database indicated the presence of pigment proteins. In

233   particular, RPE65, which functions in the cleavage of carotenoids, was found. In *H. axyridis*, increased

234   carotenoid pigmentation has been linked to increased alkaloid amounts (Britton et al., 2008). In addition,

235   the elytral coloration of the seven spot ladybug, *Coccinella septempunctata*, is a result of several

236   carotenoids (Britton et al., 2008). While larva are mostly black (**Figure 1a**), we posit that the orange

237   sections on the lower back could be due to carotenoid production. Moreover, this study provides a

238   necessary foundation for the continued study of the genetic link between genes and the maintenance

239   of variation in *H. axyridis.*

240

241

242   **Literature Cited:**

243   Bezzerides A. L., McGraw K. J., Parker R. S., Husseini J. (2007). Elytral color as a signal of chemical
244   defense in the Asian ladybird beetle *Harmonia axyridis*. *Behavior Ecology and Sociobiology* 61:
245   1401-1408.

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

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

250   Bray N., Pimental H., Melsted P., Lior P. (2015). Near-optimal RNA-seq quantification. eprint
251   arXiv:1505.02710

252   Britton G., Liassen-Jesen, S., Pfander H.(2008). Carotenoids, Vol. 4: Natural Functions. *Springer.* DIO:
253   0.1007/978-3-7643-7499-0

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

256   Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L. (2009).
257   BLAST+: architecture and applications. *BMC Bioinformatics* doi: 10.1186/1471-2105-10-421.

258   Carbon S., Ireland A., Mungall C.J., Shu S., Marshall B., Lewis S. (2009). AmiGO Hub, Web Presence
259   Working Group. AmiGO: online access to ontology and annotation data. *Bioinformatics*
260   25(2):288-9.

261   Darwin C. (1859). The Origin of Species. New York, New York: Bantam Dell.

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

263   Endler J. A. (1986). Natural Selection in the Wild. Princeton, New Jersey: Princeton University Press.

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

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

268   Gray S. M., McKinnon J. S. (2006). Linking color polymorphism maintenance and speciation.
269   *ScienceDirect*: doi:10.1016/j.tree.2006.10.005

270   Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles
271   D., Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N.,
272   Wasterman R., William T., Dewey C. N., Henschel R., LeDuc R. D., Friedman N., Regev A.
273   (2013). *De novo* transcript sequence reconstruction from RNA-seq using Trinity platform for
274   reference generation an analysis. *Nature Protocols* 8: 1494-1512.

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

277 Jennings B. H. (2011). Drosophila – a versatile model in biology & medicine. *MaterialsToday* 14: 190-
278   195.

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

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

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

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

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

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

293 Mbandi S. K., Hesse U., Rees J. G., Christoffels A. (2014). A glance at quality score: implication for
294   de novo transcriptome reconstruction of Illumina reads. Frontiers in Genetics (5): doi:
295   10.3389/fgene.2014.00017

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

298 Patro R., Suggal G.,Kingsford C. (2015). Accurate, fast, and model-aware transcript expression
299   quantification with Salmon. bioRxiv: **doi:** http://dx.doi.org/10.1101/021592

300 Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V., Zdobnov E. M. (2015). BUSCO:
301   assessing genome assembly and annotation completeness with single-copy orthologs.
302   *Bioinformatics* dio: 10.1093/bioinformatics/btv351.

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

306 Wheeler T. J., Eddy S. R. (2013). nhmmer:DNA homology search with profile HMMs. (2013).
307   *Bioinformatics*: doi: 10.1093/bioinformatics/btt403