Characterizing the adult and larval transcriptome of the multicolored Asian lady beetle, *Harmonia axyridis*

Lindsay A Havens, Matthew D MacManes

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*.

1 **Characterizing the Adult and Larval Transcriptome of the Multicolored Asian Lady Beetle,**

2 ***Harmonia axyridis***

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

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

8 * lao95@wildcats.unh.edu

9 + matthew.macmanes@unh.edu

13 ***Abstract:***

14 The reasons for the evolution and maintenance of striking visual phenotypes are as widespread as the

15 species that display these phenotypes. While study systems such as *Heliconius,* the neotropic butterfly,

16 and *Dendrobatidae,* poison dart frogs, have been well characterized and provide critical information

17 about the evolution of these traits, a breadth of new study systems, in which the phenotype of interest

18 can be easily manipulated and quantified, are essential for gaining a more general understanding of

19 these specific evolutionary processes. One such model is the multicolored Asian lady beetle, *Harmonia*

20 *axyridis*, which displays significant elytral spot and color polymorphism. Using transcriptome data from

21 two life stages, adult and larva, we characterize the transcriptome, thereby laying a foundation for further

22 analysis and identification of the genes responsible for the continual maintenance of spot variation in

23 *H. axyridis*.

24

25

26

27

28

29

30

31

32 ***Introduction:***

33 The evolution and maintenance of phenotypic polymorphism and striking visual phenotypes have

34 fascinated scientists for many years (Darwin, 1859; Endler, 1986; Fisher, 1930; Gray and McKinnon,

35 2006; Joron et al., 2006). In general, insects have become increasingly popular as study organisms to

36 examine phenotypic variation (Jennings, 2011; Joron et al., 2006). One such insect displaying extensive

37 elytra and spot variation that has yet to be extensively studied is the Asian Multicolored Ladybeetle,

38 *Harmonia axyridis*.

39

40 The mechanisms responsible for the evolution of these phenotypes are as widespread as the species

41 that display them. Aposematism, crypsis, and mimicry may play a role in the evolution of phenotypic

42 variation in the animal kingdom. Members of family *Dendrobatidae,* poison dart frogs, are

43 aposematically colored (Cadwell, 1996), while *Tetrix subulata* grasshoppers maintain their phenotypic

44 polymorphism to aid in crypsis (Karpestam et al., 2014). A mimicry strategy is utilized by one particularly

45 well-characterized species that exhibits phenotypic polymorphism, the Neotropic butterfly system,

46 *Heliconius*. The color, pattern, and eyespot polymorphism seen in *Heliconius* is thought to have arose

47 as a result of Müllerian mimicry (Flanagan et al., 2004) and the supergenes underlying these traits have

48 been well characterized (Kronforst et al., 2006, Joron et al., 2006, Jones et al., 2012).

49

50 These studies, aiming to elucidate the mechanistic links between phenotype and genotype, present a

51 unique opportunity to gain insight into the inner workings of many important evolutionary processes.

52 While systems like poison frogs and butterflies have been pioneering, the use of novel models,

53 especially those that can be easily manipulated, are needed. One such study system that possesses

54 many of the benefits of classical models, while offering several key benefits, described below, is the

55 multicolored Asian lady beetle, *Harmonia axyridis*. *Harmonia*, which is common throughout North

56 America, and easily bred in laboratory environments, possesses significant variation in elytral spot

57 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 (Havens, personal observation). The patterning is symmetrical on both wings. In some

61 animals, there is a center spot beneath the pronotum which leads to an odd number of spots. The elytral

62 spots are formed by the production of melanin pigments (Bezzerides et al., 2007). The frequency of

63 different morphs varies with location and temperature (Michie et al., 2010). The melanic morph is more

64 prevalent in Asia when compared to North America (LaMana and Miller, 1996; Dobzhansky, 1933). A

65 decrease in melanic *H. axyridis* has been shown to be correlated with an increase in average yearly

66 temperatures in the 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 a 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 placed in RNAlater and immediately 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), which includes PolyA purification of mRNA. To allow multiple samples to
93 be run in one lane, a unique index code was added to each sample. These samples were then pooled
94 in equimolar quantities using Tapestation analysis. The multiplexed libraries were then sent to the New
95 York Genome Center (New York, USA) for sequencing on a single lane (125bp paired end) of the HiSeq
96 2500 sequencer.

97

98
99
100   **Figure 1a: The larva used for transcriptome sequencing.**
101

102
103
104   **Figure 1b: The adult used for transcriptome sequencing.**
105
106
107   *Sequence Data Preprocessing and Assembly*
108
109   The raw sequence reads corresponding to the two tissue types were error corrected using the software
110   BLESS (Heo et al., 2014) version 0.17 (https://goo.gl/YHxlzI, https://goo.gl/vBh7Pg). The error-
111   corrected sequence reads were adapter and quality trimmed following recommendations from
112   MacManes (MacManes, 2014) and Mbandi (Mbandi, 2014). Specifically, adapter sequence
113   contamination and low quality nucleotides (defined as Phred<2) were removed using the program

114 Trimmomatic version 0.32 (Bolger, 2014) called from within the Trinity assembler version 2.1.1 (Haas,
115 2013). Reads from each tissue were assembled together to created a joint assembly of adult and larva
116 transcripts using a Linux workstation with 64 cores and 1Tb RAM. We used flags to indicate the stranded
117 nature of sequencing reads and set the maximum allowable physical distance between read pairs to
118 999nt (https://goo.gl/ZYP08M).

120 The quality of the assembly was evaluated using transrate version 1.01 (Unna-Smith 2015;
121 (https://goo.gl/RpdQSU). Transrate generates quality statistics based on a process involving mapping
122 sequence reads back to the assembled transcripts. Transcripts supported by properly mapped reads of
123 a sufficient depth (amongst other things) are judged to be of high quality. In addition to generating quality
124 metrics, transrate produces an alternative assembly with poorly-supported transcripts removed. This
125 improved assembly was used for all downstream analyses and QC procedures. We then evaluated
126 transcriptome completeness via use of the software package BUSCO version 1.1b (Simão 2015).
127 BUSCO searches against a database of highly-conserved single-copy genes in Arthropoda
128 (https://goo.gl/bhTNdr). High quality, complete transcriptomes are hypothesized to contain the vast
129 majority of these conserved genes as they are present in most other species.

131 To remove assembly artifacts remaining after transrate optimization, we estimated transcript abundance
132 using 2 software packages - Salmon version 0.51 (Patro 2015; https://goo.gl/01UIF6) and Kallisto
133 version 0.42.4 (Bray, 2015; https://goo.gl/BsQMpr). Transcripts whose abundance exceeded 0.5 TPM
134 in either adult or larval datasets using either estimation method were retained. TPM is different from
135 FPKM with regards to the order of operations performed, as TPM normalizes for gene length first. We
136 evaluated transcriptome completeness and quality, again, after TPM filtration, using BUSCO and
137 transrate, to ensure that our filtration processes did not significantly effect the biological content of the
138 assembly.

140 We identified and removed potential plant, fungal, bacterial and vertebrate contamination by using a
141 blastx search. We created a custom protein database based on the collection of protein sequences
142 from   each   taxonomic   group   available   for   download   from   RefSeq
143 (ftp://ftp.ncbi.nlm.nih.gov/refseq/release/). We queried this database, inferring a given sequence to be
144 a contaminate if the best blast (= lowest e-value) was plant, fungal, bacterial or vertebrate on origin,
145 rather than invertebrate.

147 ***Assembled Sequence Annotation***

149 The   filtered   assemblies   were   annotated   using   the   software   package   dammit
150 (https://github.com/camillescott/dammit; https://goo.gl/05MY5i). Dammit coordinates the annotation

151 process, which involves use of blast (Camacho, 2009), TransDecoder (version 2.0.1,
152 http://transdecoder.github.io/), and hmmer version 3.1b1 (Wheeler 2013). In addition to this, putative
153 secretory proteins were identified using the software signalP, version 4.1c (https://goo.gl/FaOQSj).

154

155 In addition to this, we attempted to identify loci involved in color, color patterning, and more generally
156 phenotypic polymorphism. To accomplish this, we downloaded a set of 1008 candidate genes
157 previously identified as underlying phenotypic evolution in other species, from Dryad (Martin and
158 Orgogozo, 2013a and 2013b). We used the gene name listed in this dataset to download the protein
159 sequence from Uniref90. We used a blastX search strategy to identify potential homologues in the
160 dataset.

161

162 To identify patterns of gene expression unique to each life stage, we used the expression data as per
163 above. We identified transcripts expressed in one stage but not the other, and cases where expression
164 occurred in both life stages. The Uniprot ID was identified for each of these transcripts using a blastX
165 search (https://goo.gl/J9saMj), and these terms were used in the web interface Amigo (Carbon et al.,
166 2009) to identify Gene Ontology terms that were enriched in either adult or larva relative to the
167 background patterns of expression. The number of unique genes contained in the joint assembly was
168 calculated via a BLAST search of the complete gene sets of human, *Homo sapiens*, fruit fly, *Drosophila*
169 *melanogaster*, and the flour beetle, *Tribolium casteneda*.

170

171 **Results and Discussion:**

172

173 *Data Availability*

174

175 All read data are available under ENA accession number PRJEB13023. Assemblies and data matrices
176 are available at https://goo.gl/D3xh65, and will be moved to Dryad following manuscript acceptance.

177

178 *RNA extraction, Assembly and Evaluation*

179 RNA was extracted from whole bodies of both the adult and the larva stage of a single *Harmonia*
180 *axyridis*. The quality was verified using a Tapestation 2200 (all RIN scores > 8) as well as a Qubit. The
181 initial concentration for the larva sample was 83.2 ng*µL -1, while the initial concentration for the adult
182 sample was 74.7 ng*µL -1. The number of strand-specific paired end reads contained in the adult and
183 larva libraries were 58 million and 67 million, respectively. The reads were 125 base pairs in length.

184

185 The raw Trinity assembly of the larval and adult reads resulted in a total of 171,117 contigs (82Mb)
186 exceeding 200nt in length. 526 contaminate sequences were removed. This assembly was evaluated
187 using Transrate, producing an initial score of 0.10543, and and optimized score of 0.29729. The
188 optimized score indicated that the optimized assembly was better than 50% of NCBI-published *de novo*
189 transcriptomes (Unna-Smith 2015). This transrate optimized assembly (89,305 transcripts, 62Mb) was
190 further filtered by removing transcripts whose expression was less than 0.5 TPM. After filtration, 33,648
191 transcripts (44Mb) remained. To assess for the inadvertent loss of valid transcripts, we ran BUSCO
192 before and after this filtration procedure. The percent of Arthropoda BUSCO's missing from the
193 assembly rose slightly, from 18% to 21%. Transrate was run once again, and resulted in a final
194 assembly score of 0.29112. This score is indicative of a high-quality transcriptome appropriate for
195 further study (Unna-Smith 2015). In an attempt at understanding how many distinct genes our
196 transcriptome contained, we conducted a blast search against *Homo sapiens*, *Drosophila*
197 *melanogaster*, and *Tribolium casteneda.* This search resulted in 7,246, 7,739, and 7,741 matches
198 unique to *Harmonia axyridis*, which serve as estimates of the number of unique genes expressed in
199 these two life stages. The final assembly is available at https://goo.gl/nWdBuv

200

201 ***Annotation***

202

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

212

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

217

218 A complement of immune-related genes were discovered as well. These include a single member of
219 the Attacins and Coleoptericins, two TLR-like genes, seven Group 1, and 34 Group 2 C-type lectin

220   Receptors (CLRs). Two CARD-containing Cytoplasmic pattern recognition receptor (CRR) genes were
221   discovered, as were 3 MAP kinase containing transcripts. Finally, 119 RIG-I-like receptors (RLR) were
222   found.

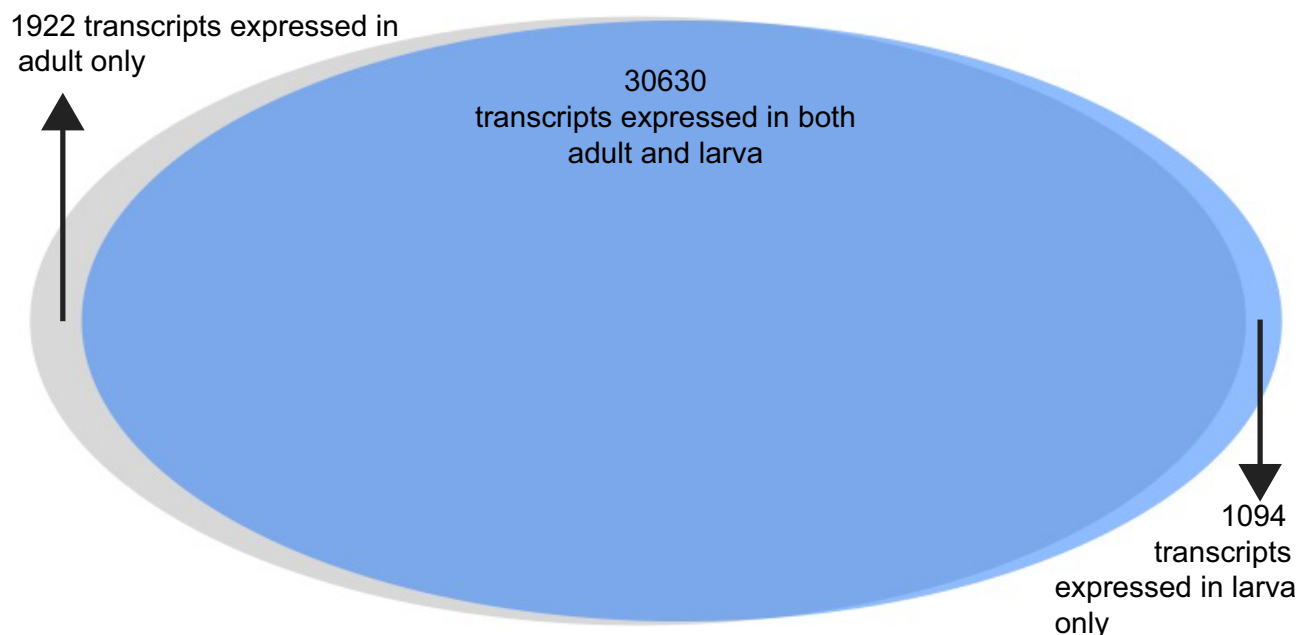
223

224   The focused search for genes previously implicated in phenotypic polymorphism (Martin and Orgogozo,
225   2013a and 2013b) identified 483 *Harmonia* transcripts, corresponding to 65 distinct loci (Supplementary
226   table 1). These loci include the transcription factors bab1, Distal-less, Optix, enzymes BCMO1, ebony,
227   yellow, and transporters ABCC2, SLC24A5, and TPCN2, all related to color and color patterning. These
228   genes, and the others identified in Supplementary table to are likely to provide fodder for future
229   research.

230

231   Gene expression was estimated for each transcript for both adult and larva (Figure 2, data available
232   here: https://goo.gl/wM3TV7). For all transcripts expressed in the adult, the mean TPM=29.7 (max=
233   80,016.9, SD= 500.8) while the mean larval TPM=29.71 (max= 166,264, SD=941). When analyzing
234   transcripts found uniquely in these two tissues, mean adult TMP=9.9 (max=3,037, SD=86) and for larva
235   TPM=8.6 (max=687, SD=41).

236

237

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

240

241   Analysis of the differences between adult and larval life stages were carried out as well. Because these
242   life stages were only sequenced with a single individual each, they should be interpreted with some

243 caution. The vast majority of transcripts were observed in both life stages (n=30,630, 91%), with a small

244 number being expressed uniquely in larva (n=1,094) and adult (n=1,922). Of these transcripts

245 expressed uniquely in either larva or adult, 45% and 42%c were annotated using at least one method

246 via the software package dammit. 6.1% and 4.6%, respectively, were found to be secretory in nature

247 via signalP analysis.

248

Conclusions:

250 Phenotypic polymorphisms and striking visual phenotypes have fascinated scientists for many years.

251 The breadth of evolutionary causes for the maintenance of these phenotypes are as numerous as the

252 species that display them. One organism, *Harmonia axyridis*, provides a unique opportunity to explore

253 the genetic basis behind the maintenance of an easy to quantify variation - elytral spot number. While

254 understanding these genomic mechanisms is beyond the scope of this paper, we do provide a reference

255 transcriptome for *H. axyridis*, a foundational resource for this work.

256

257 This study indicates that most gene expressed at levels greater than .5 TPM were shared seen in both

258 the adult and larval individuals. While the majority of proteins identified in the assembled transcriptome

259 were structural in function, analyses of protein families using the Pfam database indicated the presence

260 of pigment proteins. In particular, RPE65, which functions in the cleavage of carotenoids, was found. In

261 *H. axyridis*, increased carotenoid pigmentation has been linked to increased alkaloid amounts (Britton

262 et al., 2008). In addition, the elytral coloration of the seven spot ladybug, *Coccinella septempunctata*, is

263 a result of several carotenoids (Britton et al., 2008). While larva are mostly black (**Figure 1a**), we posit

264 that the orange sections on the lower back could be due to carotenoid production. Moreover, this study

265 provides a necessary foundation for the continued study of the genetic link between genes and the

266 maintenance of variation in *H. axyridis.*

267

268

269 **Literature Cited:**

270 Bezzerides A. L., McGraw K. J., Parker R. S., Husseini J. (2007). Elytral color as a signal of chemical
271 defense in the Asian ladybird beetle *Harmonia axyridis*. *Behavior Ecology and Sociobiology* 61:
272 1401-1408.

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

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

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

279 Britton G., Liassen-Jesen, S., Pfander H.(2008). Carotenoids, Vol. 4: Natural Functions. *Springer.* DIO:
280 0.1007/978-3-7643-7499-0

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

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

285 Carbon S., Ireland A., Mungall C.J., Shu S., Marshall B., Lewis S. (2009). AmiGO Hub, Web Presence
286 Working Group. AmiGO: online access to ontology and annotation data. *Bioinformatics*
287 25(2):288-9.

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

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

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

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

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

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

297 Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles
298 D., Li B., Lieber M., MacManes M. D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N.,
299 Wasterman R., William T., Dewey C. N., Henschel R., LeDuc R. D., Friedman N., Regev A.
300 (2013). *De novo* transcript sequence reconstruction from RNA-seq using Trinity platform for
301 reference generation an analysis. *Nature Protocols* 8: 1494-1512.

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

304 Jennings B. H. (2011). Drosophila – a versatile model in biology & medicine. *MaterialsToday* 14: 190-
305   195.

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

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

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

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

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

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

320 Mbandi S. K., Hesse U., Rees J. G., Christoffels A. (2014). A glance at quality score: implication for
321   de novo transcriptome reconstruction of Illumina reads. *Frontiers in Genetics* (5): doi:
322   10.3389/fgene.2014.00017.

323 Martin A, Orgogozo V (2013a) The loci of repeated evolution: a catalogue of genetic hotspots of
324   phenotypic variation. Evolution 67(5): 1235-1250. http://dx.doi.org/10.1111/evo.12081

325 Martin A, Orgogozo V (2013b) Data from: The loci of repeated evolution: a catalogue of genetic
326   hotspots of phenotypic variation. Dryad Digital
327   Repository. http://dx.doi.org/10.5061/dryad.v66p0

328 Michie L.J., Mallard F., Majerus M.E.N., Jiggins F.M. (2010). Melanic through nature or nurture:
329   genetic polymorphism and phenotypic plasticity in *Harmonia axyridis*. *Journal of Evolutionary
330   Biology* 23:1699-1707.

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

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

335   Simão F. A., Waterhouse R. M., Ioannidis P., Kriventseva E. V., Zdobnov E. M. (2015). BUSCO:

336   assessing genome assembly and annotation completeness with single-copy orthologs.

337   *Bioinformatics* dio: 10.1093/bioinformatics/btv351.

338   Smith-Unna R. D., Boursnell C., Patro R., Hibberd J. M., Kelly S. (2015). TransRate: reference free

339   quality assessment of *de-novo* transcriptome assemblies. *bioRxiv*: dio

340   http://dx.doi.org/10.1101/021626.

341   Wheeler T. J., Eddy S. R. (2013). nhmmer:DNA homology search with profile HMMs. (2013).

342   *Bioinformatics*: doi: 10.1093/bioinformatics/btt403