UNIVERSITY OF CALIFORNIA,
IRVINE

Diversification and expression of vision-related genes in Lepidoptera

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Ecology and Evolutionary Biology

by

Aide Macias Muñoz

Dissertation Committee:
Professor Adriana D. Briscoe, Chair
Associate Professor Kevin Thornton
Associate Professor Ali Mortazavi

2018

DEDICATION

To

my parents, Maria and Raymundo, whose sacrifices and hard work have made this possible. Your courage has inspired me to be brave in pursuing my career goals. Your strength has motivated me to work hard and your kindness has taught me to be caring. Your love and support have been monumental in my educational pursuit.

DEDICATORIA

Para

mis padres, María y Raymundo, cuyos sacrificios y trabajo duro han hecho esto posible. Su valor me ha inspirado a ser valiente en perseguir mis objetivos profesionales. Su fuerza me motiva a trabajar duro y su amabilidad me ha enseñado bondadosa. Sobre todo, su amor y apoyo han sido monumental en mi búsqueda educativa.

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	v
ACKNOWLEDGMENTS	vi
CURRICULUM VITAE	viii
ABSTRACT OF THE DISSERTATION	x
INTRODUCTION	1
CHAPTER 1: Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic	14
CHAPTER 2: Copy number variation and expression analysis reveals a non-orthologous *pinta* gene family member involved in butterfly vision	74
CHAPTER 3: Phototransduction in Lepidoptera	138
CONCLUSIONS	208

LIST OF FIGURES

Page

Figure 1.1 Photographs of adult *Bicyclus anynana* heads and heatmaps of diffentially expressed (DE) contigs by seasonal form and sex. 45

Figure 1.2 Functional enrichment of differentially expressed genes. 46

Figure 1.3 Expression of differentially expressed genes across seasonal forms. 47

Figure 1.4 Opsin expression in *B. anynana* seasonal forms. 49

Figure 1.5 Reverse-transcription PCR (RT-PCR) of eye and brain tissue. 50

Figure 1.6 Eye development gene networks. 51

Figure 2.1 Bayesian phylogeny of insect CRAL-TRIO domain proteins. 106

Figure 2.2 CRAL-TRIO domain gene location and patterns of mRNA presence or absence. 107

Figure 2.3 Expression of CRAL-TRIO domain genes. 109

Figure 2.4 *Hme CTD31* RT-PCR. 110

Figure 2.5 *Hme CTD31* Western Blot. 111

Figure 2.6 Immunohistochemistry of Hme CTD31 in H. melpomene eye and optic lobe. 112

Figure 3.1 Proposed model of the lepidopteran phototransduction cascade. 167

Figure 3.2 Differential expression analysis. 168

Figure 3.3 Opsin gene family. 169

Figure 3.4 DAGL gene family. 170

Figure 3.5 TRP gene family. 171

Figure 3.6 Wunen gene family. 172

Figure 3.7 Vha-100 gene family 173

LIST OF TABLES

		Page
Table 1.1	Summary of findings from Everett et al. (2012).	52
Table 1.2	Summary of the total number of differentially expressed (DE) contigs and unique gene ontology (GO) terms discovered in analyses.	53
Table 1.3	BLAST results for contigs with opsin-like expression patterns.	54
Table 1.4	Significant *P* values of two-factor ANOVAs for eye development, phototransduction, and eye pigment genes.	55
Table 2.1	Summary of differentially expressed (DE) and upregulated contigs.	113
Table 2.2	Expression patterns of CRAL-TRIO domain contigs.	114
Table 2.3	Top 20 proteins from upper, middle and lower bands detected by mass spectrometry sorted by upper band Mascot score.	115
Table 3.1	Summary of transcriptome-wide analysis.	174

ACKNOWLEDGMENTS

First and foremost, I would like to thank my wonderful advisor Dr. Adriana Briscoe for her support and guidance throughout my PhD. Her mentorship has been vital to my academic success and I will forever cherish the advice she has given me and all that she has taught me. She embodies what it means to be a great researcher, teacher, and mentor.

I am grateful to my committee members Dr. Kevin Thornton, Dr. Ali Mortazavi, and Dr. Francisco Ayala. Their profound knowledge of the fields of evolution, development, and genetics was invaluable in helping me determine rigorous tests for my hypotheses. Their advice during committee meetings helped me improve figures and motivated me to aim high when submitting journals. I am also grateful for the advice and mentorship provided Dr. JJ Emerson who advised me on project development and analysis. I would also like to thank my collaborator and co-author Dr. Antónia Monteiro for her intellectual contributions and advice in completing chapter 1 of my dissertation.

I would like to thank faculty, postdocs, students and staff in the Ecology and Evolutionary Department who have made working at UC Irvine a fun and stimulating experience. Their love of science and friendly demeanors have been inspiring. Dr. Kate Loudon and Dr. Donovan German always provide motivating conversation at department events and Mahul Chakraborty, Joe Heras, and Jim Baldwin-Brown are always willing to discuss molecular and bioinformatic techniques. I am thankful for Spanish nights with current and past students Alejandra Rodriguez, Will Petry, Adriana Romero, Kyle McCulloch, Alberto Soto, Wilnelia Recart, and Ana Catalán. I am also thankful for family dinners with Jimmy Kezos, Caitlin Looby, Thomas Barter, Grant Rutledge, and Mark Phillips.

I was extremely lucky to have Mark Phillips as a colleague, neighbor, and friend. Thank you for bringing me lemons when I was sick, for picking me up after midnight once when I had to stay in lab late, and for all of the cheer up dinners.

I am thankful to have had some awesome and brilliant lab members throughout the years. Kyle McCulloch, Susan Finkbeiner, Gilbert Smith, Ana Catalán, Jennifer Briner, Aline Rangel, and Zachary Johnson have given essential feedback on my presentations and writing. Gil was a colleague and a mentor; our discussions of data analysis and project design have helped shape me into the scientist that I have become. I have had the privilege to work closely with Gil, Ana and Aline to co-author publications and I admire their work ethic and value their help and input.

Thank you Kyle for making every day in lab enjoyable. Your constructive criticism and encouragement made me a better and more confident scientist. I am honored to have worked alongside you for five years. I look forward to future collaborations. I was lucky to have you as a labmate and friend. I will always treasure the memories of our salsa dancing experiences, gossip sessions, and our conference stories.

This dissertation would not be possible without the love and support of my friends and family. I was lucky to attend graduate school near my large Macias family and had the chance to see them often. Thank you all for your joy. Interacting with you reminds me of the importance of hard work. While I do not see the Muñoz family as often, I can feel your love and support, thank you for believing in me. My sister Rubi Macias is my hero and I can only hope one day to be as fierce as she is. I cannot thank my parents Raymundo and Maria Macias enough for everything that they do for my sister and me. Rubi and I owe our accomplishments to your sacrifices and lessons.

I would also like to thank my bestie Richard Garcia who makes light of everything and makes me happy every time we talk or spend time together. I want to thank Imran Khan for motivating and pushing me these past few months. I am grateful to have an amazing group of powerful women in my life. Verenice Bravo's carefree attitude encourages me to be brave and take risks. Nicole Campos is someone who I can always talk to about professional and personal problems; her wise advice has been priceless. Brittaney de la Torre has been both a source of emotional support and entertainment, thank you for tolerating my multiple musings via text every single day. Diana Gutierrez has been like a sister, thank you for providing a listening ear when I need it and for being an amazing best friend. The four of you motivate me to be successful and kind, our conversations ground me and remind me that it's important to give back to our community.

Finally, I would like to thank the funding sources that made this dissertation research possible. Fellowship support was provided by the National Science Foundation Graduate Research Fellowship Program and by the Ford Predoctoral Fellowship Program. Additional fellowship support was provided by the William D. Redfield Graduate Fellowship Award. I am grateful for travel support from the Broadening Participation Committee Travel Award, Charlotte Mangum Student Support, UC Irvine Associated Graduate Student Travel Grant, BEACON Center Travel Award and support from the Society for the Study of Evolution.

CURRICULUM VITAE

Aide Macias Muñoz

EDUCATION

2018 University of California Irvine, CA
Ph.D. in Ecology and Evolutionary Biology
2015 University of California Irvine, CA
M.S. in Ecology and Evolutionary Biology
2011 University of California Berkeley, CA
B.A. in Integrative Biology

PUBLICATIONS

Smith G, Kelly JE, **Macias-Muñoz A**, Butts CT, Martin RW, Briscoe AD. 2018. Evolutionary and structural analyses uncover the role of novel gene duplications in pollen feeding butterflies. *Proceedings of the Royal Society of London B*

Macias-Muñoz A, McCulloch KJ, Briscoe AD. 2017. Copy number variation and expression analysis reveals a non-orthologous *pinta* gene family member involved in butterfly vision. *Genome Biology and Evolution* 9:3398-3412.

Smith G, **Macias-Muñoz A**, Briscoe AD. 2016. Gene duplication and gene expression changes play a role in the evolution of candidate pollen feeding genes in *Heliconius* butterflies. *Genome Biology and Evolution* 8:2581-2596.

Macias-Muñoz A, Smith G, Monteiro A, Briscoe AD. 2016. Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic. *Molecular Biology and Evolution* 33:79–92.

Smith G, **Macias-Muñoz A**, Briscoe AD. 2014. Complete genome sequence of a novel Iflavirus from mRNA sequencing of the butterfly *Heliconius erato*. *Genome Announc*. 2: e00398-14.

Briscoe AD, **Macias-Muñoz A**, Kozak KM, Walters JR, Yuan F, et al. 2013. Female behaviour drives expression and evolution of gustatory receptors in butterflies. *PLoS Genetics* 9: e1003620.

APPOINTMENTS

September- December 2012	Teaching Assistant Human Physiology
January-March 2013	Teaching Assistant Organisms to Ecosystems
April-June 2013	Teaching Assistant Honors Physiology
April-June 2013	Teaching Assistant Population Dynamics

SELECTED AWARDS

2014 William D. Redfield Graduate Fellowship Award
2013 Ford Foundation Predoctoral Fellowship
2013 NSF Graduate Research Fellowship

SELECTED SERVICE

2017 Co-founder of a SACNAS chapter at UC Irvine
2017 Reviewer for SSE Rosemary Grant Award
2016 Society for the Study of Evolution Graduate Student Advisory Council member
2016 Reviewer for SSE Rosemary Grant Award
2015 Reviewer for *BMC Evolutionary Biology*

ABSTRACT OF THE DISSERTATION

Diversification and expression of vision-related genes in Lepidoptera

By

Aide Macias Muñoz

Doctor of Philosophy in Ecology and Evolutionary Biology

University of California, Irvine, 2018

Professor Adriana D. Briscoe, Chair

Diversification of lepidopteran (moth and butterfly) color vision is due to gene duplication, gene loss, and sequence variation of opsin genes. How other vision-related genes are evolving in this group is not well characterized. This dissertation aims to survey the evolution and expression of candidate genes involved in lepidopteran vision. In my first study, I tested whether vision-related genes varied in expression between sexes and seasonal forms of a phenotypically plastic butterfly, *Bicyclus anynana*. *B. anynana* displays plasticity and sexual dimorphism in eye size. I identified eye development genes differentially expressed between seasonal forms, making them candidates underlying eye size differences. I found that more genes were differentially expressed between seasonal forms than sexes, with rearing temperature having a larger effect on the expression of vision-related genes in females. One of the genes that was differentially expressed between seasonal forms was annotated to have a CRAL-TRIO domain. Members of this gene family have a role in chromophore transport in vertebrates and *Drosophila* (*pinta*). This gene family has been shown to be evolving by linage-specific duplications in insects and has an expansion in Lepidoptera. For my next study, I used phylogenetics and transcriptomics to

assess the evolution and expression patterns of members of this gene family in the butterfly *Heliconius melpomene*. Results indicated that a family member non-orthologous to *Drosophila pinta* takes on the role of chromophore binding. This brought upon the question: which phototransduction genes are conserved between Lepidoptera and *Drosophila*? For the last study of my dissertation, I used phylogenetics and transcriptomics to explore more gene families in order to identify which phototransduction cascade genes were conserved between *Drosophila* and Lepidoptera, and within Lepidoptera. Results suggested that moth and butterfly phototransduction cascades involve similar genes that vary from *Drosophila* cascades in instances where Lepidopteran-specific paralogs have a potential role in vision. This dissertation explored the role of plasticity in vision-related gene expression (chapter 1), provided the first evidence of visual function for a CRAL-TRIO domain gene in butterflies (chapter 2), and offered insights into the Lepidopteran phototransduction cascade (chapter 3).

INTRODUCTION

One of the main questions in the field of visual ecology is: how are animal visual systems evolving and specializing in diverse environments? The insect order Lepidoptera, which consists of moths and butterflies, provides an interesting group to investigate this question because Lepidoptera are active during variable light conditions. Lepidoptera is the second largest insect order, containing around 160,000 known species of butterflies and moths (Wahlberg et al. 2013; Kawahara & Breinholt 2014). Studies suggest that butterflies are a monophyletic clade whereas moths are paraphyletic (Regier et al. 2013; Kawahara & Breinholt 2014; Mitter et al. 2017). Moth and butterfly visual systems are morphologically and functionally adapted to their light environments, with some exceptions. Most butterflies are active during the day or twilight and have apposition type eyes (Frederiksen & Warrant 2008). Apposition eye type is a compound eye structure common to diurnal insects that allows a photon to be processed by each separate unit of the eye (ommatidium) (Nilsson et al. 1984). Conversely, moths are characteristically nocturnal with superposition eyes adapted for low light environments. Superposition compound eyes have a space between the crystalline cone and the light-processing rhabdom so that each rhabdom can collect photons from multiple facet lenses (Warrant & Dacke 2016). However, some exceptions exist such as diurnal moths (Kelber et al. 2003; Feuda et al. 2016), nocturnal butterflies (Yack & Fullard 2000; Yack et al. 2007), and diurnal butterflies with superposition eyes (Horridge et al. 1972). This visual diversity within the group propagates the question: how are lepidopteran visual systems evolving?

Most studies on lepidopteran vision focus on investigating color vision, specifically by focusing on opsin genes. Butterflies use color vision to forage for food, choose mates,

and identify host plants (Kelber 1999; Jiggins et al. 2001; Robertson & Monteiro 2005; Snell-Rood & Papaj 2009; Finkbeiner et al. 2014; Kinoshita & Arikawa 2014). Although nocturnal, some moths have been shown to have color vision (Cutler et al. 1995; Kelber et al. 2002, 2003; Xu et al. 2013). Opsin genes encode a protein that binds a chromophore, a vitamin-A derived molecule, to form a light sensitive pigment called rhodopsin. Photon absorption causes a change in the configuration of the chromophore from 11-*cis* to all-*trans* initiating the phototransduction cascade. The basic lepidopteran eye has three opsins used for color vision: blue (B), ultraviolet (UV) and long-wavelength (LW) (Stavenga & Arikawa 2006; Briscoe 2008; Xu et al. 2013; Feuda et al. 2016). Lepidopteran opsins are evolving by gene duplications, gene losses, and changes in amino acid sequence. Opsin duplication events facilitate the increasing complexity of color vision by providing material that can expand animal spectral sensitivities. Some examples in butterflies include LW duplications that enable seeing red (Briscoe 2001; Arikawa et al. 2003; Frentiu et al. 2007; Chen et al. 2016), a blue duplication to see into green (Sison-Mangus et al. 2008), and a UV duplication to see in the violet range (McCulloch et al. 2016).

Dynamic opsin gene evolution is not unique to Lepidoptera. Some beetle species that have lost a blue opsin gene show spectral sensitivity in the blue range due to changes in the coding sequence of UV and LW duplications (Sharkey et al. 2017). In addition, a survey of 10 dragonfly families uncovered variation in opsin number between families ranging from 11 to 30 visual opsins across species (Futahashi et al. 2015). Outside of insects, opsins in ray-finned fish have expanded by tandem duplications and have diversified by varying selective pressures of different light environments (Seehausen et al. 2008; Rennison et al. 2012). In primates, a long wavelength opsin duplication is

responsible for trichromatic vision (Dulai et al. 1999; Surridge et al. 2003). While investigating the dynamic evolution of opsins gives insight into visual system evolution, the molecular evolution and function of other genes involved in vision, such as phototransduction, should also be considered (Plachetzki & Oakley 2007; Plachetzki et al. 2007). Rivera et al. (2010) found that pancrustaceans have high rates of retention and duplication of vision-related genes. These studies suggest that visual systems across species might vary in their signaling pathway due to gene duplications and changes in protein interactions. *This dissertation aims to 1) explore the role of plasticity in vision-related gene expression, 2) identify a chromophore-binding protein in butterflies, and 3) probe the conservation of the phototransduction cascade between Drosophila and Lepidoptera.*

Role of plasticity

Phenotypic plasticity refers the ability of a genotype to create different phenotypes in response to environmental fluctuations (Schlichting & Smith 2002). Plasticity is maintained when environmental change is predictable and each phenotype has a higher fitness in their respective environments (Moran 1992). In addition to developmental plasticity, organisms can display phenotypic flexibility which is a change in phenotype to acclimate to an environmental change. When it comes to vision, studies in fish have found plasticity in opsin expression in cichlids reared in light environments lacking a UV component (Hofmann et al. 2010). A recent study also found that medaka fish exhibit plasticity of phototransduction genes leading to seasonal changes in behavior and light perception (Shimmura et al. 2017). In the housefly, *Musca domestica*, plasticity of phototransduction allows males to have a faster voltage response specialized for chasing

behavior (Hornstein et al. 2000). Plasticity in opsin gene expression is observed in a female-specific manner in the phenotypically plastic butterfly *Bicyclus anynana* (Everett et al. 2012).

Wild populations of the butterfly *B. anynana* encounter a wet and dry season to which they respond to by variation in wing coloration (Brakefield & Reitsma 1991). In the lab, the wet and dry season morphs can be triggered by changing their rearing temperature (Kooi & Brakefield 1999). In addition to phenotypic plasticity, *B. anynana* display rex-role reversal in the dry season where the females court males (Bear & Monteiro 2013). Accompanying these variable phenotypic and behavioral traits, sensory systems are also plastic in this butterfly species. Eye size is sexually dimorphic and varies within sexes between seasonal forms and opsin expression is decreases in non-choosy dry season females (Everett et al. 2012). These observations raise the question: which genes are involved in sensory plasticity in *B. anynana*?

In the first chapter of my dissertation, I hypothesized that genes involved in eye development, eye pigmentation and phototransduction would vary between sexes and seasonal forms. I tested my hypothesis by generating RNA-Sequencing libraries from whole heads of wet and dry season males and females then performed differential expression analyses. I found that there was a larger effect of season rather than sex on head transcription and that season had a larger effect on females. The vision-related genes that were differentially expressed between seasonal forms had functions related to eye development and eye pigmentation. The results from this study suggest that difference in expression of eye development genes might be driving divergent eye phenotypes and that female visual systems are more plastic in this species (Macias-Muñoz et al. 2016).

Chromophore-binding

One of the genes that was down-regulated in *B. anynana* dry season females was annotated with an alpha-tocopherol transport function and a functional enrichment of genes differentially expressed by season revealed a cluster annotated for cellular-retinaldehyde binding. Cellular retinaldehyde binding protein (CRALBP) in vertebrates functions in visual chromophore 11-*cis*-retinal transport (Wu et al. 2006). Chromophores are derived from vitamin A and are hydrophobic so require transport by a specialized protein (Panagabko et al. 2003). In *Drosophila*, the chromophore 11-*cis*-3-hydroxyretinal is transported by a CRAL-TRIO domain containing protein encoded by prolonged depolarization afterpotential (PDA) is not apparent (*pinta*) (Wang & Montell 2005). A survey of insect CRAL-TRIO domain genes found that this gene family is evolving by lineage-specific duplications and that Lepidoptera have twice as many copies relative to other species sampled (Smith & Briscoe 2015). This study generated a phylogeny which showed that Lepidoptera did not have a *pinta* ortholog (Wang & Montell 2005; Smith & Briscoe 2015).

In Chapter 2 of my dissertation I aimed to identify a CRAL-TRIO domain gene involved in butterfly chromophore transport. I hypothesized that one or more members of this gene family would be upregulated in butterfly heads implying a role in vision. In order to test my hypothesis, I generated RNA-Seq libraries from heads, antennae, legs and mouth parts of the butterfly *Heliconius melpomene*. I used *H. melpomene* because this butterfly species has a reference genome and resequenced genome data is publicly available for *H. melpomene* subspecies (Martin et al. 2013; Davey et al. 2016). Genome data allowed me to explore the molecular evolution of the CRAL-TRIO domain gene family in one additional

butterfly species not included in the previous insect survey (Smith & Briscoe 2015). By searching the genome and a *de novo* transcriptome, I found 43 CRAL-TRIO domain genes in *H. melpomene* and a species-specific expansion. In addition, I found that many of the genes are located in tandem and some have potential copy number variation among individuals, especially of the genes found within the expansion.

Differential expression analyses revealed that one gene, *Hme CTD31*, was upregulated in heads across comparisons to antennae, legs, and mouth parts. This gene is present as a single copy in the 18 resequenced genomes that I investigated. To validate that this gene was expressed in the eye and not the brain, I did RT-PCR of RNA from eye and brain tissue and found *Hme CTD31* expressed in only the eye. To localize the gene product, I did immunohistochemistry on head sections that included retina and brain. The protein Hme CTD31 was expressed in primary pigment cells, secondary pigment cells, and tracheal cells. The results of this study supported evidence that the CRAL-TRIO domain gene family is evolving by lineage specific duplications. In addition, I found a function in butterfly vision for a member of the gene family that is rapidly evolving in Lepidoptera.

Phototransduction in insects

The CRAL-TRIO domain gene that encodes a protein which binds the chromophore in *H. melpomene*, *Hme CTD31*, is not a *Drosophila pinta* ortholog. This brings into question which phototransduction genes are conserved between *Drosophila* and Lepidoptera, especially since phototransduction genes are evolving by duplication events (Rivera et al. 2010). Investigating the conservation of the phototransduction cascade in moths and butterflies is important because the only insect in which genes have been experimentally tested for a role in vision is *Drosophila* (Hardie & Raghu 2001; Montell 2012; Hardie &

Juusola 2015). Other insect studies that explore the expression of phototransduction genes infer their function based on sequence similarity to *Drosophila* (French et al. 2015). For my third chapter, I used transcriptomics and phylogenetics to survey the molecular evolution and expression of lepidopteran genes annotated with *Drosophila* phototransduction genes. I hypothesized that were would be Lepidopteran-specific duplications. Within Lepidoptera, moths and butterflies are active in different light environments and have different compound eye structures to accommodate for these differences in light availability. I expected to see differences in gene loss/gain or gene expression within Lepidoptera, between moths and butterflies underlying the variation in visual systems.

Genes associated with vision and phototransduction are highly expressed in butterfly heads which are composed of mostly eye and optic lobe (Montgomery et al. 2016). To verify that genes expressed in heads functioned in vision, I used the gene comparisons from Chapter 2 and annotated differentially expressed genes between heads, legs, antennae, and mouth parts in *Heliconius melpomene*. I was particularly interested in the genes upregulated in heads across comparisons. Annotation of the commonly upregulated genes revealed that these genes function in light detection and regulation of the rhodopsin signaling pathway. The genes upregulated in butterfly heads were potentially *Drosophila* homologs which would suggest a conserved function of genes across species, and conservation of the phototransduction cascade. To validate this and identify any potential gene gains or losses between species, we generated genetic phylogenies for 31 gene families, searching genomes of *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Tribolium casteum*, *Bombyx mori*, *Manduca sexta*, *Danaus plexippus* and *Heliconius melpomene* for 57 genes.

Phylogenetic trees revealed that most genes are conserved as single copies between *Drosophila* and Lepidoptera. However, I identified 2 gene losses in Lepidoptera and one gene loss in non-lepidopteran insects. I used RNA-Seq data from *M. sexta* heads and *H. melpomene* heads, legs, antennae, and mouth parts to infer function in vision by quantifying expression of homologs in eyes. I found instances where a Lepidopteran-specific gene is involved in vision and one occurrence of *Drosophila* maintaining one gene paralog and Lepidoptera maintaining another, but outgroup insects maintained both. Within Lepidoptera, there was also no consistent difference between moths and butterflies in vision gene gain or loss. Yet, I did find a Lepidopteran-specific duplication of a gene involved in compound eye development. These results suggest that we can use *Drosophila* as a model for insect phototransduction but the molecular evolution and expression of these genes should be investigated to determine gene conservation. To validate a role in vision, future studies need to turn off each of the candidates and test photoreceptor response to light.

Overall my dissertation finds that plasticity and molecular evolution play a role in the diversification of lepidopteran vision. Previous studies have focused on describing opsin content and testing for color vision, these are the first studies done to expand on the opsins and look at the rest of the phototransduction cascade while exploring the role of phenotypic plasticity, gene duplications, and sexual dimorphism on lepidopteran visual systems. These results provide insight into how visual transduction is evolving within and between clades.

REFERENCES

Arikawa K, Mizuno S, Kinoshita M, Stavenga DG. 2003. Coexpression of two visual pigments in a photoreceptor causes an abnormally broad spectral sensitivity in the eye of the butterfly *Papilio xuthus*. J. Neurosci. 23:4527–4532.

Bear A, Monteiro A. 2013. Male courtship rate plasticity in the butterfly *Bicyclus anynana* is controlled by temperature experienced during the pupal and adult stages. PLoS One. 8:e64061.

Brakefield PM, Reitsma N. 1991. Phenotypic plasticity, seasonal climate and the population biology of *Bicyclus* butterflies (Satyridae) in Malawi. Ecol. Entomol. 16:291–303.

Briscoe a D. 2001. Functional diversification of lepidopteran opsins following gene duplication. Mol. Biol. Evol. 18:2270–2279.

Briscoe AD. 2008. Reconstructing the ancestral butterfly eye: focus on the opsins. J. Exp. Biol. 211:1805–13.

Chen P, Awata H, Matsushita A, Yang E, Arikawa K. 2016. Extreme spectral richness in the eye of the common bluebottle butterfly , *Graphium sarpedon*. Front. Ecol. Environ. 4:18. doi: 10.3389/fevo.2016.00018.

Cutler D, Bennett R, Stevenson R, White R. 1995. Feeding behavior in the nocturnal moth *Manduca sexta* is mediated mainly by blue receptors, but where are they located in the retina? J. Exp. Biol. 198:1909–17.

Davey JW et al. 2016. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. G3. 6:695–708.

Dulai KS, Von Dornum M, Mollon JD, Hunt DM. 1999. The evolution of trichromatic color vision by opsin gene duplication in new world and old world primates. Genome Res. 9:629–638. doi: 10.1101/gr.9.7.629.

Everett A, Tong X, Briscoe AD, Monteiro A. 2012. Phenotypic plasticity in opsin expression in a butterfly compound eye complements sex role reversal. BMC Evol. Biol. 12:232. doi: 10.1186/1471-2148-12-232.

Feuda R, Marle F, Bentley MA, Holland PWH. 2016. Conservation, duplication, and divergence of five opsin genes in insect evolution. Genome Biol. Evol. 8:579–587. doi: 10.5287/bod-leian.

Finkbeiner SD, Briscoe AD, Reed RD. 2014. Warning signals are seductive: Relative contributions of color and pattern to predator avoidance and mate attraction in *Heliconius* butterflies. Evolution (N. Y). 68:3410–3420. doi: 10.1111/evo.12524.

Frederiksen R, Warrant EJ. 2008. Visual sensitivity in the crepuscular owl butterfly *Caligo memnon* and the diurnal blue morpho *Morpho peleides*: a clue to explain the evolution of nocturnal apposition eyes? J. Exp. Biol. 211:844–851. doi: 10.1242/jeb.012179.

French AS, Meisner S, Liu H, Weckström M, Torkkeli PH. 2015. Transcriptome analysis and RNA interference of cockroach phototransduction indicate three opsins and suggest a major role for TRPL channels. Front. Physiol. 6:207. doi: 10.3389/fphys.2015.00207.

Frentiu FD, Bernard GD, Sison-Mangus MP, Van Zandt Brower A, Briscoe AD. 2007. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. Mol. Biol. Evol. 24:2016–2028. doi: 10.1093/molbev/msm132.

Futahashi R et al. 2015. Extraordinary diversity of visual opsin genes in dragonflies. Proc. Natl. Acad. Sci. 112:E1247–E1256. doi: 10.1073/pnas.1424670112.

Hardie RC, Juusola M. 2015. Phototransduction in *Drosophila*. Curr. Opin. Neurobiol. 34:37–45. doi: 10.1016/j.conb.2015.01.008.

Hardie RC, Raghu P. 2001. Visual transduction in *Drosophila*. Nature. 413:186–193. doi: 10.1038/35093002.

Hofmann CM, O'Quin KE, Smith AR, Carleton KL. 2010. Plasticity of opsin gene expression in cichlids from Lake Malawi. Mol. Ecol. 19:2064–2074. doi: 10.1111/j.1365-294X.2010.04621.x.

Hornstein EP, O'Carroll DC, Anderson JC, Laughlin SB. 2000. Sexual dimorphism matches photoreceptor performance to behavioural requirements. Proc. R. Soc. B Biol. Sci. 267:2111–2117. doi: 10.1098/rspb.2000.1257.

Horridge GA, Giddings C, Stange G. 1972. The superposition eye of skipper butterflies. Proc. R. Soc. London. Ser. B, Biol. Sci. 182:457–495. doi: 10.1098/rspb.1972.0088.

Jiggins CD, Naisbit RE, Coe RL, Mallet J. 2001. Reproductive isolation caused by colour pattern mimicry. Nature. 411:302–305. doi: 10.1038/35077075.

Kawahara AY, Breinholt JW. 2014. Phylogenomics provides strong evidence for relationships of butterflies and moths. Proc. R. Soc. B Biol. Sci. 281:20140970. doi: 10.1098/rspb.2014.0970.

Kelber A. 1999. Ovipositing butterflies use a red receptor to see green. J. Exp. Biol. 202:2619–2630. papers://4b986d00-906f-493f-a74b-71e29d82b719/Paper/p1.

Kelber A, Balkenius A, Warrant EJ. 2003. Colour vision in diurnal and nocturnal hawkmoths. Integr. Comp. Biol. 43:571–579.

Kelber A, Balkenius A, Warrant EJ. 2002. Scotopic colour vision in nocturnal hawkmoths. Nature. 419:922–925. doi: 10.1038/nature01065.

Kinoshita M, Arikawa K. 2014. Color and polarization vision in foraging *Papilio*. J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol. 200:513–526. doi: 10.1007/s00359-014-0903-5.

Kooi RE, Brakefield PM. 1999. The critical period for wing pattern induction in the polyphenic tropical butterfly *Bicyclus anynana* (Satyrinae). J. Insect Physiol. 45:201–212.

Macias-Muñoz A, Smith G, Monteiro A, Briscoe AD. 2016. Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic. Mol. Biol. Evol. 33:79–92. doi: 10.1093/molbev/msv197.

Martin SH et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23:1817–1828. doi: 10.1101/gr.159426.113.

McCulloch KJ, Osorio D, Briscoe AD. 2016. Sexual dimorphism in the compound eye of *Heliconius erato*: a nymphalid butterfly with at least five spectral classes of photoreceptor. J. Exp. Biol. 219:2377–2387. doi: 10.1242/jeb.136523.

Mitter C, Davis DR, Cummings MP. 2017. Phylogeny and evolution of Lepidoptera. Annu. Rev. Entomol. 62:265–283. doi: 10.1146/annurev-ento-031616-035125.

Montell C. 2012. *Drosophila* visual transduction. Trends Neurosci. 35:356–363. doi: 10.1016/j.tins.2012.03.004.Drosophila.

Montgomery SH, Merrill RM, Ott SR. 2016. Brain composition in *Heliconius* butterflies, posteclosion growth and experience-dependent neuropil plasticity. J. Comp. Neurol. 524:1747–1769. doi: 10.1002/cne.23993.

Moran NA. 1992. The evolutionary maintenance of alternative phenotypes. Am. Nat. 139:971–989.

Nilsson DE, Land MF, Howard J. 1984. Afocal apposition optics in butterfly eyes. Nature. 312:561–563. doi: 10.1038/312561a0.

Panagabko C et al. 2003. Ligand specificity in the CRAL-TRIO protein family. Biochemistry. 42:6467–6474. doi: 10.1021/bi034086v.

Plachetzki DC, Degnan BM, Oakley TH. 2007. The origins of novel protein interactions during animal opsin evolution. PLoS One. 2. doi: 10.1371/journal.pone.0001054.

Plachetzki DC, Oakley TH. 2007. Key transitions during the evolution of animal phototransduction: Novelty, 'tree-thinking,' co-option, and co-duplication. Integr. Comp. Biol. 47:759–769. doi: 10.1093/icb/icm050.

Regier JC et al. 2013. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). PLoS One. 8:e58568. doi: 10.1371/journal.pone.0058568.

Rennison DJ, Owens GL, Taylor JS. 2012. Opsin gene duplication and divergence in ray-finned fish. Mol. Phylogenet. Evol. 62:986–1008. doi: 10.1016/j.ympev.2011.11.030.

Rivera AS et al. 2010. Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. BMC Evol. Biol. 10:123. doi: 10.1186/1471-2148-10-123.

Robertson KA, Monteiro A. 2005. Female *Bicyclus anynana* butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. Proc. R. Soc. B Biol. Sci. 272:1541–6.

Schlichting C, Smith H. 2002. Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. Evol. Ecol. 16:189–211.

Seehausen O et al. 2008. Speciation through sensory drive in cichlid fish. Nature. 455:620–626. doi: 10.1038/nature07285.

Sharkey CR et al. 2017. Overcoming the loss of blue sensitivity through opsin duplication in the largest animal group, beetles. Sci. Rep. 7:1–10. doi: 10.1038/s41598-017-00061-7.

Shimmura T et al. 2017. Dynamic plasticity in phototransduction regulates seasonal changes in color perception. Nat. Commun. 8:1–7. doi: 10.1038/s41467-017-00432-8.

Sison-Mangus MP, Briscoe AD, Zaccardi G, Knuttel H, Kelber A. 2008. The lycaenid butterfly *Polyommatus icarus* uses a duplicated blue opsin to see green. J. Exp. Biol. 211:361–369. doi: 10.1242/jeb.012617.

Smith G, Briscoe AD. 2015. Molecular evolution and expression of the CRAL_TRIO protein family in insects. Insect Biochem. Mol. Biol. 62:168–173. doi: 10.1016/j.ibmb.2015.02.003.

Snell-Rood EC, Papaj DR. 2009. Patterns of phenotypic plasticity in common and rare environments: A study of host use and color learning in the cabbage white butterfly *Pieris rapae*. Am. Nat. 173:615–631. doi: 10.1086/597609.

Stavenga DG, Arikawa K. 2006. Evolution of color and vision of butterflies. Arthropod Struct. Dev. 35:307–318. doi: 10.1016/j.asd.2006.08.011.

Surridge AK, Osorio D, Mundy NI. 2003. Evolution and selection of trichromatic vision in primates. Trends Ecol. Evol. 18:198–205. doi: 10.1016/S0169-5347(03)00012-0.

Wahlberg N, Wheat CW, Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS One. 8:1–8. doi: 10.1371/journal.pone.0080875.

Wang T, Montell C. 2005. Rhodopsin formation in *Drosophila* is dependent on the PINTA retinoid-binding protein. J. Neurosci. 25:5187–94.

Warrant E, Dacke M. 2016. Visual navigation in nocturnal insects. Physiology. 31:182–192. doi: 10.1152/physiol.00046.2015.

Wu Z et al. 2006. CRALBP ligand and protein interactions. Adv. Exp. Med. Biol. 572:477–483.

Xu P et al. 2013. The evolution and expression of the moth visual opsin family. PLoS One. 8:e78140. doi: 10.1371/journal.pone.0078140.

Yack JE, Fullard JH. 2000. Ultrasonic hearing in nocturnal butterflies. Nature. 403:265–266. doi: 10.1038/35002247.

Yack JE, Johnson SE, Brown SG, Warrant EJ. 2007. The eyes of *Macrosoma* sp. (Lepidoptera: Hedyloidea): A nocturnal butterfly with superposition optics. Arthropod Struct. Dev. 36:11–22. doi: 10.1016/j.asd.2006.07.001.

CHAPTER 1

Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic

ABSTRACT

Vision is energetically costly to maintain. Consequently, over time many cave-adapted species down-regulate the expression of vision genes or even lose their eyes and associated eye genes entirely. Alternatively, organisms that live in fluctuating environments, with different requirements for vision at different times, may evolve phenotypic plasticity for expression of vision genes. Here we use a global transcriptomic and candidate gene approach to compare gene expression in the heads of a polyphenic butterfly. *Bicyclus anynana* have two seasonal forms that display sexual dimorphism and plasticity in eye morphology, and female-specific plasticity in opsin gene expression. Non-choosy dry season females down-regulate opsin expression, consistent with the high physiological cost of vision. To identify other genes associated with sexually dimorphic and seasonally plastic differences in vision we analyzed RNA-Sequencing data from whole head tissues. We identified two eye development genes (*klarsicht* and *warts* homologs) and an eye pigment biosynthesis gene (*henna*) differentially expressed between seasonal forms. By comparing sex-specific expression across seasonal forms, we found that *klarsicht*, *warts*, *henna*, and another eye development gene (*domeless*) were plastic in a female-specific manner. In a male-only analysis, *white* (*w*) was differentially expressed between seasonal forms. RT-PCR confirmed that *warts* and *white* are expressed in eyes only, whereas *klarsicht, henna* and *domeless* are expressed in both eyes and brain. We find that differential expression of eye development and eye pigment genes is associated with

divergent eye phenotypes in *B. anynana* seasonal forms, and that there is a larger effect of season on female vision-related genes.

INTRODUCTION

Eyes are metabolically expensive tissues. A rare study of the energetic requirements of photoreceptor cells found that up to 8% of a fly's resting metabolic rate was consumed by their eyes (Niven 2014). Given the high cost of vision, it is unsurprising that when organisms colonize new environments with low to no light their eyes often degenerate, presumably in order to free up energetic resources for re-allocation to other tissues (Fong et al. 1995). As an example, flies kept in captivity have smaller eyes due to a reduction in facet number, which reduces photoreceptor energy consumption (Tan et al. 2005). Another more extreme example is the colonization of caves and underground habitats that generally leads to eye reduction or loss in a broad range of species through parallel evolutionary changes in key genes contributing to visual atrophy (Culver and Pipan 2009). In many of these cave-adapted animals, changes in the expression of eye development, phototransduction, and eye pigment genes have been associated with eye size reduction or loss. A transcriptomic analysis of a cave-adapted beetle, *Ptomaphagus hirtus*, showed the absence of some eye pigment and photoreceptor genes in head transcriptomes to be correlated with reduced compound eyes (Friedrich et al. 2011; Friedrich 2013). Similarly, in the fish genus *Sinocyclocheilus*, photoreceptor genes were down-regulated in a cavefish species relative to a surface species (Meng et al. 2013). The expression of developmental genes, which can either repress or promote gene expression depending on developmental context, can be more complex. For instance, the expression of *hedgehog*, was significantly

lower in the eyes of the cave amphipod *Gammarus minus*, relative to surface species (Aspiras et al. 2012), whereas in the Mexican blind cavefish, *Astyanax mexicanus*, higher expression of a *hedgehog* ortholog was found to drive eye degeneration (Yamamoto et al. 2004). Although extreme examples of eye loss suggest fixed and potentially irreversible genetic changes (pseudogenizations or even gene deletions) that enhance fitness by reducing investment in vision, phenotypic plasticity in visual systems is a relatively unexplored form of adaptation (likely involving more subtle changes in the regulation of gene expression levels) that may also be an evolutionarily important mechanism for coping with the high cost of vision.

Plasticity in eye morphology can indicate phenotypic plasticity in vision, and may be accompanied by plasticity in expression of vision-related genes. Developmental phenotypic plasticity is the ability of a single genotype to create different (fixed) phenotypes in response to environmental cues that are usually experienced early in development of the organism or even earlier in the maternal environment (Schlichting and Smith 2002). Adaptive plasticity evolves when populations live in recurrent fluctuating environments producing different phenotypes in each that have a higher fitness in their respective environments (Moran 1992). Plasticity can evolve and play a role in evolutionary processes such as adaptation and speciation. For example, plasticity evolves through standing genetic variation or *de novo* mutation resulting in canalized traits or changes in plasticity to optimize fitness (Nijhout 2003; West-Eberhard 2003; Pigliucci et al. 2006; Crispo 2007). Phenotypic plasticity can thus affect the probability and direction of genetic evolution and may drive diversification (Price et al. 2003; Aubin-Horth and Renn 2009; Pfennig et al. 2010). Additionally, plasticity may play a role in adaptation through a direct influence on

reproductive isolation and the promotion of evolutionary responses by nonadaptive plasticity (Fitzpatrick 2012). Furthermore, while some plastic traits in an individual can be adaptive, other plastic traits accompanying them may be maladaptive (Steinger et al. 2003; Ghalambor et al. 2007). Developmental phenotypic plasticity is distinct from phenotypic flexibility, where individuals are flexible and can acclimate to changing environments. An example of visual flexibility has been found in fish, where the type of opsin expressed in an individual fish retina will shift as the fish ages (Hofmann et al. 2010). Another example of physiological changes in vision is the circadian regulation of opsin expression levels linked to diurnal light-dark cycles (Sasagawa et al. 2003; Spaethe and Briscoe 2005; Battelle et al. 2013). Here we explore phenotypic plasticity in vision of *Bicyclus anynana* butterflies whose distinct cohorts live in recurrent fluctuating environments in Africa, consisting of a dry season followed by a wet season (Brakefield and Reitsma 1991), and whose alternative vision phenotypes have been proposed to be adaptive (Everett et al. 2012).

Bicyclus anynana is a plastic sex-role reversed species that exhibits phenotypic plasticity in wing pattern morphology and sexual behavior. The wet season (WS) form has conspicuous eyespots and a pale band on its wings, while the dry season (DS) form has cryptic coloration and reduced eyespots (Brakefield and Reitsma 1991; Brakefield et al. 2009; Monteiro et al. 2015). In the laboratory, the WS and DS season wing forms are induced by rearing butterflies at 27°C and 17°C respectively (Koch et al. 1996; Kooi and Brakefield 1999). *B. anynana'*s courtship behavior is controlled by temperature experienced during pupal development and early adulthood (Prudic et al. 2011; Bear and Monteiro 2013). In conditions resembling the WS (27°C), males court females but in the DS (17°C), females court males (Prudic et al. 2011). In addition, wet season females (WSF) and

dry season males (DSM) exhibit choosy behavior. Females choose mates based on UV-reflectance of the dorsal forewing eyespot pupils (Robertson and Monteiro 2005); choosy males (DSM) and choosy females (WSF) prefer mates with intact UV pupils (Prudic et al. 2011). UV-reflectance brightness in these ornaments is highest in WSM, followed by DSF, WSF, and finally DSM (Everett et al. 2012). WSM also have larger eyespot centers compared to DSM, while there is no difference between female seasonal forms (Prudic et al. 2011).

Everett et al. (2012) posited that non-choosy individuals would have relaxed selection for vision accompanied by plasticity in eye morphology or visual sensitivity due to energetic costs of maintaining enhanced vision (Niven et al. 2007; Niven and Laughlin 2008; Niven 2014). This hypothesis was partially upheld by data from female *B. anynana* but not males. Eye size measurements demonstrated that eye size is both sexually dimorphic and phenotypically plastic; males had larger eyes compared to females and WS forms had larger eyes in both sexes compared to DS forms (Figure 1.1A)(Table 1.1; Everett et al. 2012). Facet number was also greater in males compared to females and in WS forms compared to DS, while facet size was larger in males and DS forms compared to females and WS forms, respectively (Table 1.1; Everett et al. 2012). Most significantly for the current study, qPCR showed that *B*. *anynana* long-wavelength (*LWRh*), blue (*BRh)*, and ultraviolet (*UVRh)* opsin genes have decreased expression in non-choosy DS female butterflies, but not in males (Everett et al. 2012). The plasticity and sexual dimorphism of *B. anynana* eye phenotypes makes this a suitable system with which to: 1) identify additional vision-related genes on a whole transcriptome level, which to our knowledge has not been done in butterflies before, and 2) investigate how variation in expression levels of these genes is correlated with previously measured phenotypic differences. We

hypothesized that eye developmental pathway genes would be differentially expressed between individuals with different eye sizes (i.e., males vs. females; dry vs. wet season). We also hypothesized that since opsins were differentially expressed in DS and WS females, additional phototransduction genes would be down-regulated in non-choosy DSF with smaller eyes and lower opsin expression. Lastly, to relate our findings in butterflies to eye development in other organisms, we examined our transcriptomes for the presence or absence of genes in several signaling pathways that may be conserved across arthropods: visual system specification, retinal determination, and photoreceptor differentiation (Rivera et al. 2010).

We used a high-throughput RNA-Sequencing (RNA-Seq) approach to examine differential expression between *B. anynana* sexes and seasonal forms in order to identify genes associated with divergent eye phenotypes (Wang et al. 2009; Colombo et al. 2013; Meng et al. 2013; Smith et al. 2013). We explored differences between DSF, DSM, WSF and WSM and, in order to validate larger effects of seasonal form on females, we compared effects of seasonal forms within sexes. A large number of genes were differentially expressed (DE) between seasonal forms and sexes, including 3 vision-related genes (*klarsicht, warts*, and *henna*) that were DE across seasonal forms. These 3 genes in addition to *domeless* were differentially expressed across seasonal forms within females only, while a single vision-related gene (*white*) was DE across male forms. A vision-related (eye development, phototransduction, and eye pigment) candidate gene approach showed that 36 and 18 genes had P values < 0.05 for season and sex contrasts, respectively. We find that plasticity in expression of eye development and eye pigment genes is associated with

divergent eye phenotypes in *B. anynana* seasonal forms, and that season has a larger effect on female visual systems.

RESULTS AND DISCUSSION

Whole transcriptome expression patterns

RNA-Seq libraries were constructed using mRNA extracted from 0-3 hour old adult whole head tissue (excluding mouth parts and antennae) of 12 *B. anynana* individuals; 3 biological replicates of each of the 4 specimen types: 3 DSF, 3 WSF, 3 DSM and 3 WSM. In total, we sequenced 12 libraries using high-throughput Illumina sequencing producing 100 bp paired-end reads. Quality trimming resulted in approximately 12 million reads per trimmed library. Several *de novo* assembly protocols were explored using Trinity (Grabherr et al. 2011), and the final reference assembly consisted of 43,248 contigs with an N50 of 2,299 bp. On average, approximately 85% of reads were successfully aligned to the assembly across libraries (Table S1.1).

To identify differentially expressed contigs between the treatment types, we fit a generalized linear model (glm) in edgeR (Robinson et al. 2010) with terms for sex, seasonal form, and a sex×seasonal form interaction on raw count data from all 12 libraries. Using a false discovery rate threshold (FDR) of < 0.05, we found 722 contigs that were differentially expressed across seasonal forms, 290 across sexes, and 111 showed a significant interaction between sex and seasonal form (Table 1.2). Heatmaps for DE genes across seasonal forms (Figure 1.1B) and sexes (Figure 1.1C) showed clear groupings of gene expression for each factor and their interaction (Figure S1.1). We used BLAST+ (Camacho et al. 2009) to identify homologous genes in *Drosophila* and assign gene ontology (GO) terms to our contigs (Table S1.2). We found 229 unique GO terms corresponding to

DE contigs for seasonal form, 77 for sex, and 23 for the interaction of seasonal form and sex (Table 1.2).

DE genes potentially associated with differences in eye size

Everett et al. (2012) found that wet season eyes are generally larger than dry season eyes and that male eyes are larger than female eyes (Table 1.1). Functional enrichment tests were performed using DAVID (Huang et al. 2009) for each model term (sex, seasonal form, and interaction) to group contigs with similar annotation terms into functional clusters, in order to identify genes associated with differences in eye size. Enrichment of DE contigs by seasonal form produced the highest number of enriched clusters (Figure 1.2A-F, Table S1.3). Among these, we found two gene clusters, which may reflect decreased eye size in DS forms. Annotation cluster 1 included contigs annotated with insect cuticle protein structure; nine of these contigs were homologous to named cuticular protein genes. In cluster 1, 10 of 15 contigs were up-regulated in DS forms (Figure 1.2A). In cluster 3, 13 of 21 contigs up-regulated in DS forms had functions involving extracellular regions, aminoglycan and chitin metabolic process (Figure 1.2C). During larval development, a second stage of head tissue cell fate commitment consists of differentiation into retina or head cuticle (Friedrich 2003). These gene expression patterns may reflect a larger number of head cuticle secreting cells in DS forms. Annotation cluster 2, enriched for cellular retinaldehyde and alpha-tocopherol transfer (Figure 1.2B), may have a more direct effect on vision if these contigs retain a similar function to that of related gene family member *pinta*, a gene that encodes a visual chromophore binding and transport protein in *Drosophila* (see below).

Enrichment results for DE contigs across sexes resulted in four annotation clusters of genes encoding extracellular region proteins, immunoglobulin, cell adhesion, and calcium ion binding proteins (Figure 1.2G-J, Table S1.3). 14 of the 16 enriched DE contigs for sex were up-regulated in females. Enrichment of DE contigs showing an interaction between sex and seasonal forms resulted in one significant cluster of contigs encoding endopeptidase activity proteins (Table S1.3). Overall, a functional enrichment of our DE contigs identified only two annotation clusters that may reflect differences observed in eye size. However, with the possible exception of annotation cluster 2 for seasonal DE contigs, enrichment tests did not detect specific vision-related clusters so we manually inspected GO terms associated with each of our DE contigs.

Within DE contigs across seasonal forms we found two contigs homologous to *Drosophila* genes that influence eye development, and an eye pigment biosynthesis gene (Table S1.4). *Warts* (*wts*) determines opsin expression in R8 photoreceptor cells in *Drosophila* that in turn differentiates pale from yellow ommatidia types, crucial to color discrimination (Mikeladze-Dvali et al. 2005). *Wts* was up-regulated in DSF compared to DSM and WS forms (Figure 1.3A). *Klarsicht* (*klar*) affects eye morphology and *klar* mutants in *Drosophila* have a rough eye phenotype driven by malformed photoreceptors (Mosley-Bishop et al. 1999); the contig homologous to this gene was down-regulated in DS forms (Figure 1.3B). *Henna* (*Hn*) is an eye pigment biosynthesis gene (GO:0006726) and was down-regulated in DSF (Figure 1.3C). We did not find any vision-related genes in lists of DE contigs across sexes or displaying an interaction between seasonal forms and sex (Table S1.4). Although the two-factor analysis identified these three genes as being differentially expressed between seasons, close visual inspection of their plotted FPKMs (fragments per

kilobase of exon per million fragments mapped) (Figure 1.3) revealed that it is the DS female form that is primarily responsible for the biggest magnitude change observed.

Female-specific analysis

As mentioned above, both the qPCR findings of Everett et al. (2012) (Table 1.1) and now our two-factor DE analysis suggest that female gene expression shows the largest plasticity. To further explore female-specific differential expression, we performed a single factor comparison of the two seasonal forms for females only. We found 790 DE contigs across seasonal forms (FDR < 0.05; Table 1.2, Table S1.4), 555 of which overlapped with DE contigs across seasonal forms in the two-factor model (Figure S1.2). A heatmap of female DE contigs across seasonal forms showed a clear grouping between seasonal forms with approximately two-thirds of contigs being up-regulated in DSF (Figure S1.3). A functional enrichment analysis for 267 (Table 1.2) unique *Drosophila* homologs resulted in 13 annotation clusters (Table S1.3) of which there was considerable functional and gene overlap with annotations clusters 1-5 of the seasonal DE genes shown in Figure 1.2. We visually inspected our list of female-specific DE contigs (Table S1.4) for vision-related GO terms and confirmed that *klar* and *Hn* were down-regulated in DSF and *wts* was up-regulated in DSF (Figure 1.3A-C). In addition to recovering these genes, we found an eye development gene differentially expressed between female seasonal forms. *Domeless* (*dome*) is a target in the JAK/STAT pathway that regulates compound eye size and morphogenesis (Tsai and Sun 2004) and was down-regulated in DSF (Figure 1.3D). Taken together, *Klar, wts* and *dome,* are candidates for investigation of their potential role in driving eye size differences. Their role in photoreceptor differentiation and eye

morphogenesis may contribute to fewer facets in DSF, and thus their DE may contribute to a smaller eye phenotype (Everett et al. 2012).

Role of sex combs reduced in *B. anynana* eye morphology

Developmental genes, such as transcription factors, often play important roles in trait development because they directly regulate the expression of other genes. In *Drosophila*, the hox gene *Sex combs reduced* (*Scr*) has a sex differentiation function (GO:0007548) and has been suggested to act in sex-specific differentiation of the basitarsus tissue (Sánchez and Guerrero 2001) and to control the development of the sex-combs in male T1 legs (Tanaka et al. 2011). Here, we hypothesized that WSF display masculine expression patterns of hox gene expression during photoreceptor differentiation causing male-like eye morphology. We inspected our female-specific DE contigs for *Scr* (Table S1.4) and found it to be differentially expressed between seasonal forms. *Scr* expression was similar in WSF, DSM, and WSM while it was up-regulated in DSF. A one-way ANOVA comparison of FPKM between WSF and males confirmed that *Scr* expression did not significantly vary between these groups (DSM P=0.52; WSM P=0.51). *Scr* has been found in the maxillary and labial palps in the head segments of several insect embryos (Rogers et al. 1997; Kokubo et al. 1997; Passalacqua et al. 2010), but no detailed knowledge of its later expression domain is known for Lepidoptera. It is possible that *Scr* could be differentially regulating the expression of genes in these head segments of DSF relative to the other three forms. An eye-specific expression for *Scr* is currently not known for any insect.

In order to reinforce our inspection of female-specific differences between seasons, we performed differential gene expression analysis between male seasonal forms. 359 contigs were differentially expressed within males, corresponding to 96 unique *Drosophila*

genes. An enrichment of these homologs resulted in one cluster enriched for phagocytosis (Table S1.3). Manual inspection of this list uncovered only one vision-related gene, *white* (Table S1.4). *White* (*w*) functions in compound eye pigmentation (GO:0048072) and was up-regulated in DSM (Figure 1.3E). In addition, we searched male DE contigs for sex differentiation GO terms and did not find any matches making *Scr* a good candidate for masculinization of eyes in females.

Candidate gene approach

Opsins and eye pigment-related genes are down-regulated in DSFs

Although Everett et al. (2012) observed differences in eye size between seasonal forms and between sexes, our whole transcriptome DE analyses yielded only a handful of candidate vision-related genes. This may be because the threshold for detecting a significant log-fold difference using this method is too high. Nonetheless, we were interested in whether our RNA-Seq data could confirm the qPCR results of Everett et al. (2012) for the opsins. Opsin genes are the core component of visual systems because they encode proteins that bind a light-absorbing chromophore and together comprise the visual pigment rhodopsin. Rhodopsin initiates the phototransduction cascade and its absorption spectrum determines photoreceptor cell sensitivity (Briscoe and Chittka 2001). The chromophore in butterflies is 11-*cis* 3-hydroxy retinal. Everett et al. (2012) found that opsin genes have decreased expression in *B. anynana* non-choosy DSF relative to choosy WSF (Table 1.1). It was previously hypothesized that choosy individuals should have enhanced vision to detect sexual ornament brightness (dorsal eyespot centers) and non-choosy individuals should have diminished vision due to physiological costs (Everett et al. 2012). We therefore expected to see a higher sensitivity to light, especially in the UV range,

for WSF because they choose mates based on the UV-reflectance of their white centers (Robertson and Monteiro 2005). Analyses of variance (ANOVAs) of mRNA expression levels quantified by calculating FPKM, normalized within and between libraries, were used to examine opsin expression levels between treatments.

In general, non-choosy DSF have decreased expression relative to the other three groups, which validates prior qPCR results from Everett *et al.* (2012). The long-wavelength (*LWRh*) opsin gene was the third most highly expressed gene across libraries, however *LWRh* opsin expression was not significantly different between sexes or seasonal forms nor was the interaction between these factors significant (Figure 1.4A, sex: F = 0.203 *P* = 0.664, seasonal form: F = 0.010 *P* = 0.992, sex*seasonal form: F = 3.888 *P* = 0.084). The blue (*BRh*) opsin gene was on average the eighteenth most highly expressed gene and was differentially expressed between seasonal forms (Figure 1.4B, sex: F = 1.570 *P* = 0.246, seasonal form: F = 10.211 *P* = 0.013, sex*seasonal form: F = 3.416 *P* = 0.102). A TukeyHSD test showed that this difference came from comparing DSF to WSF and WSM. One-way ANOVAs confirmed this trend in DSF relative to WSF (*P* = 0.029) and WSM (*P* = 0.027), but not DSM (*P* = 0.112). The ultraviolet (*UVRh*) opsin was the fifty-fifth most expressed gene across libraries, and did not show differential expression between groups using a two-way ANOVA (Figure 1.4C, sex: F = 0.041 *P* = 0.845, seasonal form: F = 3.014 *P* = 0.121, sex*seasonal form: F = 3.945 *P* = 0.082). However, one-way ANOVAs showed that DSF had decreased expression relative to WSF (*P* = 0.012), but not relative to males in either season (DSM *P* = 0.305, WSM *P* = 0.1042). Furthermore, WSM and DSM had similar levels of expression in all three opsin genes (LW *P* = 0.345, blue *P* = 0.361, UV *P* = 0.897). Our results showed that non-choosy DSF do indeed down-regulate *BRh* and *UVRh* mRNA relative to

WSF, suggesting a decreased sensitivity. Down-regulation of opsin genes was not observed in non-choosy males, which may be due to differences in energetic demands between the sexes. Females have the additional metabolic burden of producing eggs so may be under greater selective pressure to reduce non-essential physiological functions.

We hypothesized that additional phototransduction or eye pigmentation genes might be regulated in a similar manner to the opsins. Since similar expression patterns provide insight into functional categories (Eisen et al. 1998), we searched our female-specific DE genes for contigs with opsin-like patterns of expression and explored their putative functions. We found 102 contigs with log fold change (logFC) patterns of expression that were similar to the opsins, but eliminated 80 that were too variable after plotting their FPKM values and visually inspecting them. We used blastx against NCBI to determine the functions of the remaining 22 contigs and found functional descriptions for 16 of these genes through comparisons to other insects (Table 1.3). Using this approach, we identified two vision-related contigs in particular that had expression patterns similar to the opsins. One contig potentially involved in phototransduction encoded a protein homologous to alpha-tocopherol transport protein (Figure 1.3F, *comp44923*), which may be important for vision if it has a similar function to the gene *pinta* (*prolonged depolarization afterpotential is not apparent*). The *Drosophila* gene *pinta* is found in retinal pigment cells and preferentially binds all-trans-retinol *in vitro* (Wang and Montell 2005). The contig we identified, similar to *pinta*, has a CRAL/TRIO domain and is orthologous to *M. sexta* Msex010502, which is part of a large protein family that has undergone expansion and lineage-specific duplications in lepidopterans (Smith and Briscoe 2015). Another contig was a *henna*-like transcript (*Hn*) that regulates eye pigment biosynthesis (Bel et al.

1992); this contig was also differentially expressed using the whole-transcriptome two-factor and one-factor comparisons. Many, though not all butterfly photoreceptors have filtering pigments that affect light sensitivity and color vision (Briscoe 2008). Loss of eye filtering pigments has been reported in some butterflies that may rely more on chemosensory modalities, specifically olfaction and gustation for foraging and mate choice, rather than on vision (Briscoe and Bernard 2005). It is possible that the observed down-regulation of an eye pigmentation gene in our phenotypically plastic species may, under more extreme environments, be followed by loss of entire pigmentation-related pathways.

To validate the vision-related function of these DE genes in *Bicyclus*, we examined their expression in eye tissue. We did reverse transcription polymerase chain reactions (RT-PCRs) for *wts*, *klar*, *Hn*, *dome*, *w* and *comp44923* using eye (retina + optic lobe) and brain (without optic lobe) tissue. We found that *wts*, *w* and *comp44923* are only expressed in eye tissue (Figure 1.5). *Klar*, *Hn*, *dome* and positive control *BRh* are expressed in both eye and brain but have seemingly higher expression in eyes (Figure 1.5).

Butterfly eye development, phototransduction, and eye pigment genes

Since butterfly head tissue is primarily composed of eyes (retina + optic lobe), we expected to find a large number of vision-related genes in our assembly, even if few were identified as being significantly DE in the transcriptome-wide analysis. Since this study also represents, to our knowledge, the first transcriptome-wide characterization of candidate vision genes in butterflies, we undertook a manual search of the Trinity assembly for additional candidate genes involved in eye development, phototransduction, and pigmentation (Hardie 2001; Kumar 2001; Jeffery 2005; Friedrich et al. 2011). We found 203 genes and tested each gene for differential expression using two-way ANOVAs on

FPKM values. False Discovery Rate (FDR) corrections to *P* values from sex, season, and interaction effects were applied, after which only one gene was found to be significantly differentially expressed (Table S1.5). Although most of these vision-related genes were not significantly DE after multiple testing corrections, it is interesting to note the top significant genes based on uncorrected *P* values (Table 1.4). We found 34 genes involved in eye development with $P < 0.05$ across effects (e.g., *csw*, *dac*, *Egfr*, *eya*, *klar*, *toe*, and *toy*). Thirteen of these genes were DE across sex, 27 across seasonal form, and 5 showed a significant interaction. We found 8 phototransduction genes with $P < 0.05$ for sex, seasonal form, and/or interaction. Three phototransduction genes were DE across sex (*cry*, *rdgA, shakB*), 4 across seasonal form (*Pld*, *rdgC*, *BRh, shakB*) and 3 showed a significant interaction (*Arr2*, *CG11426, shakB*). Lastly, 8 eye pigment genes had $P < 0.05$ for sex, seasonal form and/or interaction. Three genes were DE across sexes (*p*, *st, w*), five across seasonal form (*dor*, *Dysb*, *lt, or, w*), and one showed a significant interaction (*Pu)*.

Eye development gene networks in butterflies

To relate our candidate vision genes in butterflies to known eye development networks in other arthropods, we inspected them for homologs of *Drosophila* genes involved in visual system specification, retinal determination, and photoreceptor differentiation. We first examined a gene network controlling visual system specification, *wingless/*Armadillo *(wg/Arm*) (Rivera et al. 2010). We identified homologs of *split ends* (*spen*), *wingless* (*wg*) and *armadillo* (*arm*) in our assembly but not *eyeless* (*ey*), *spi, rho* or *CycE*. *Spen* is a positive regulator of the *wg/Arm* signaling pathway that controls a variety of cellular processes during development (Chang et al. 2008). *Spen* expression in the developing eye stimulates *wg/Arm* signaling and causes a small eye phenotype in

Drosophila (Chang et al. 2008). We explored the expression of these genes and found no difference in expression between different sexes and seasonal forms (Figure 1.6A). One reason for the absence of several genes in this network from our transcriptome as well as this lack of difference in expression could be that we sampled gene expression after eye development and growth were complete – in early emerging adults (see Das Gupta et al. 2015).

The next gene network we explored was one controlling retinal determination (Rivera et al. 2010). *Hedgehog* (*hh*) signaling acts upstream of *decapentaplegic* (*dpp*), which affects regulatory genes *eyeless* (*ey*), *eyes absent* (*eya*), *sine oculis* (*so*), and *dachshund* (*dac*; Pappu et al. 2003). The regulatory proteins encoded by *ey*, *eya*, *so*, and *dac* are critical for retinal determination and eye development (Pappu et al. 2003). *Hn* and *dpp* initiate eye morphogenesis and, together with *eya*, are required for the progression of the morphogenetic furrow; *dac* is required for the initiation of the furrow, but not its progression, and induced *dac* expression can cause ectopic eye development (Chen et al. 1997). We found homologs of *dpp*, *eya*, *so*, and *dac* in our *Bicyclus de novo* assembly but we did not find an *ey* ortholog (Figure 1.6B), given that this gene stops being expressed in adult heads of *B. anynana* (Das Gupta et al. 2015). Two-way ANOVAs of these genes show that *dac* expression varies ($P < 0.05$) between seasonal forms and *eya* varies by sex ($P < 0.05$) (Table 1.4), while there is no difference in expression for *so* nor *dpp*. Because *dac* and *eya* function to induce eye development, we predicted that their expression would be up-regulated in specimens with larger eyes. We found that *eya* fits this predicted pattern (higher expression in males with larger eyes) whereas the opposite trend was true for *dac*, which had higher expression in DS forms that have smaller eyes relative to WS forms.

Lastly, we explored genes involved in photoreceptor differentiation which involves several interacting developmental pathways. We explored genes involved in the signaling of the homeobox, Notch, JAK/STAT (Janus Kinase/Signal Transducer and Activator of Transcription), and EGFR (epidermal growth factor receptor) signaling pathways. Firstly, we found a *BarH1* (*B-H1*) homolog in our *Bicyclus* transcriptome. In *Drosophila*, *B-H1* is a homeobox gene necessary for R1 and R6 photoreceptor progenitor differentiation and primary pigment cell development (Higashijima et al. 1992). For the Notch pathway, we found a *Notch* homolog that was not differentially expressed (Figure 1.6C, Table S1.5), however, *Aftiphilin* (*Afti*) expression varied ($P < 0.05$) between seasonal forms (Table 1.4) and had higher expression in WS forms with larger eyes. *Aftiphilin* modulates the Notch pathway and a knockdown of this gene in *Drosophila* results in irregular ommatidial size and neuronal disruption (Kametaka et al. 2012). While JAK/STAT genes *hopscotch* (*hop*) and *Signal-transducer and activator of transcription protein at 92E* (*Stat92E*) were not significantly differentially expressed, we found their target *domeless* (*dome*) up-regulated in WSF (Figure 1.6D). Furthermore, in *Drosophila*, *Egfr* plays a critical role in R8 spacing (Baonza et al. 2001) during the morphogenetic furrow which stimulates the differentiation of R8 cells (Freeman 1997). An *Egfr* homolog had higher expression in WS forms ($P < 0.05$; Table 1.4, Figure 1.6E).

We expected to find eye developmental genes differentially expressed between males and females and between seasonal forms because males have larger eyes relative to females and WS forms have larger eyes relative to DS forms (Everett et al. 2012). The expression pattern for these genes should match their effect on eye size phenotype (i.e., genes that cause “small eye phenotypes” when induced should be up-regulated in small eye

individuals). We found that some developmental genes followed expected expression patterns (e.g., *eya*, *dome*, *Egfr*), while others did not (e.g., *dac*, *wg*, *hh*, *N*, *hop*, *Stat92E*). However, we note that many genes within a developmental pathway directly affect or modulate a cell's response to another pathway (Figure 1.6). For example, *spen* (wg/arm regulator) stimulates EGFR signaling, and EGFR and JAK/STAT pathways are antagonistic to Notch (Frankfort and Mardon 2004; Doroquez et al. 2007; Flaherty et al. 2009). While our transcriptome-wide analysis identified just a handful of DE genes, our ANOVA results suggest that key eye developmental pathway genes vary in expression between seasonal forms and sexes and could be driving divergent phenotypes, especially differences in eye morphology. Up-regulation of eye photoreceptor differentiation genes in individuals with larger eyes may coincide with a higher facet number in these individuals (Everett et al. 2012).

Eye loss and the evolution of phenotypic plasticity

The possible role of developmental phenotypic plasticity in shaping vision is not well documented. Previous studies of cave-adapted animals examined presumably fixed genetic differences that contributed to eye reduction or eye loss. We expect loss of vision to be accompanied by consistent down-regulation of phototransduction genes or their absence in transcriptomes due to accumulated mutations or pseudogenization (Lahti et al. 2009; Friedrich et al 2011). In visual plasticity, we expect vision-related genes to maintain their coding sequences but vary in expression in predictable patterns determined by environmental conditions. Vision loss studies have compared the transcriptomes of different species (Friedrich et al. 2011; Meng et al. 2013) or different populations (Aspiras et al. 2012) which may have substantially diverged at the genomic level, yet these studies

do provide a starting list of candidate genes to explore in the context of visual plasticity. In this study we compared transcripts from individuals from the same stock population merely reared at different temperatures after egg laying. Although we did not create inbred lines to ensure genetic similarity between the individuals used because of the difficulty of doing so in butterflies, the observed differences are nonetheless likely due to phenotypic plasticity because our results--at least for the opsins--have now been replicated twice in the current study and in Everett et al. (2012).

CONCLUSIONS

Previous studies have correlated vision gene presence or absence or expression differences with extreme eye phenotypes, while the molecular basis of phenotypically plastic changes in eye morphology and physiology remained obscure. In our study we combined analysis of whole transcriptomes with a candidate gene approach to identify differentially expressed genes potentially driving divergent phenotypes in a polyphenic butterfly. We found that opsin genes (*BRh* and *UVRh*), a pigment biosynthesis (*henna*), and a possible eye pigment transport gene (*comp44923*) were down-regulated in non-choosy, DSF relative to WSF. Moreover, we found 3 eye development genes (*klarsicht, warts*, and *domeless*) differentially expressed between DSF and WSF that might contribute to smaller eyes in DSF. Lastly, we identified genes in developmental signaling pathways that varied in expression between sexes and seasonal forms. We propose that genes regulating developmental pathways are good candidates for driving divergent eye phenotypes, however, future studies should sample transcriptomes earlier during development to better capture differences during important differentiation stages. Our results suggest that plasticity of vision-related gene expression, particularly in females, may underlie eye

phenotypic variation and this plasticity in visual systems is likely to be of evolutionary importance.

MATERIALS AND METHODS

Animals and RNA extraction

Butterflies were reared at Yale University and at the National University of Singapore at 17°C and 27°C to produce the dry and wet season forms respectively, see Everett *et al.* (2012) for full husbandry details. Adults were frozen at -80°C on the morning of emergence when only ~ 0-3 hours old. The butterflies were shipped to UC Irvine on dry ice and stored at -80°C until RNA extraction. RNA was extracted using TRIzol (Life Technologies, Grand Island, NY) from the heads of 12 individual animals; 3 dry season females (DSF), 3 wet season females (WSF), 3 dry season males (DSM), and 3 wet season males (WSM). RNA was DNase-treated and purified using a NucleoSpin RNA II kit (Macherey-Nagel, Bethlehem, PA). Purified RNA was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY) and quality checked using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). A TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, CA) was used to make 12 double stranded cDNA libraries from our polyadenylated RNA. A Qubit Fluorometer and an Agilent Bioanalyzer were used to quantify and quality check the libraries after preparation. Libraries were then normalized and pooled according to their concentrations. Pooled libraries were run on a 2% agrose gel and size selected for DNA at ~280-340 bp. A Geneclean III kit (MP Biomedical, Santa Ana, CA) was used to recover and purify DNA from the gel, and Agencourt AMPure XP (Beckman Coulter, Brea, CA) beads were used for a second

purification. Libraries were sequenced in the UCI Genomics High-Throughput Facility using a HiSeq 2500 (Illumina, San Diego, CA), paired end 100-cycle sequence run.

Assembly and read-mapping

Raw sequenced reads were low-quality trimmed and parsed using custom perl and python scripts. *De novo* transcriptome assemblies were constructed using Trinity (Grabherr et al. 2011; Haas et al. 2013). Several different assembly protocols were considered, such as using trimmed or untrimmed libraries and varying the number of genotypes used to construct the assembly. The final 'reference assembly' chosen, based on low contig number and longest N50, was constructed using 4 trimmed libraries from 1 of each treatment type (DSF, WSF, DSM, WSM). Each sequenced library was then mapped back to the reference assembly using RSEM (Li and Dewey 2011) from which we extracted raw read count data and FPKM (fragments per kilobase of exon per million fragments mapped). FPKM was further normalized between libraries using NOISeq (Tarazona et al. 2011) to compare expression of candidate vision genes.

Whole transcriptome analysis

We performed differential gene expression analysis for all Trinity assembled contigs using edgeR, a Bioconductor package that uses a variety of statistical models to analyze read count data (Robinson and Smyth 2007, 2008; Robinson et al. 2010, McCarthy et al. 2012). A generalized linear model was fit to raw count data from all 12 libraries; this model included terms for sex, seasonal form, and sex×seasonal form interaction. We also fit a model to each of the sexes (6 libraries) comparing seasonal effects, in order to examine season-dependent gene expression within the two sexes separately. These analyses included filtering to remove contigs expressed at less than 1 count per million (cpm) for at

least 3 groups, and between sample normalization using a trimmed mean of the log expression ratios (TMM) (Robinson and Oshlack 2010). Contigs were considered significantly differentially expressed when the false discovery rate (FDR) was less than 0.05 (Storey et al. 2003; Dabney 2014). Results were visualized by creating heatmaps of differentially expressed genes using the Heatplus R package (Ploner 2012).

Opsin genes

For preliminary analysis, we used CLC Genomics Workbench (CLC bio) to create a *de novo* assembly of our libraries. We used long-wavelength, blue, and ultraviolet (*LWRh*, *BRh*, *UVRh*) opsin sequences from *Danaus plexippus* (monarch) and *Bombyx mori* (silk moth) to extract matching *B. anynana* sequences. We used MEGA 5 (Tamura et al. 2011) alignments to determine consensus sequences for each opsin coding gene. Sequences for the three opsin genes were aligned to our Trinity assembly using command-line Basic Local Alignment Search Tool (BLAST+; Camacho et al. 2009) to identify the contig ID of the top match. We plotted normalized FPKM for these contigs in R (R Core Team 2013) to visualize expression levels. For each contig, we performed two-way ANOVAs and one-way ANOVAs to compare treatment groups.

Opsin-like patterns of contig expression

To identify contigs with similar expression patterns as the opsins, we searched our female-specific DE contig list for contigs down-regulated in DSF. We retained contigs with a positive log fold change (logFC), similar to that of the opsin genes. Since opsin genes had small logFC between male seasonal forms, we reduced our list further by eliminating contigs with large logFC between male seasonal forms. We confirmed expression pattern by plotting FPKM and eliminating contigs whose expression did not resemble that of the

opsins through visual inspection. Functions of remaining contigs were determined by a nucleotide BLAST, blastx, against the NCBI database (The Uniprot Consortium).

Gene Ontology terms

We used TransDecoder in the Trinity suite to extract protein coding transcripts and amino acid sequences from our Trinity assembly. BLAST+ was used to align peptide sequences to *Drosophila* Flybase (Marygold et al. 2013) translated sequences. The top best hit for each sequence was retained when the E-value was less than 1×10^{-5}. Gene ontology (GO) terms for homologous proteins were obtained from Flybase. Functional enrichment analyses of DE contigs were performed using a Database for Annotation, Visualization, and Integrated Discovery (DAVID; Huang et al. 2009) v6.7, which grouped genes with similar functions into functional clusters.

Candidate genes

We searched *Drosophila* homologs for genes involved in eye development, adaptation to dark conditions, and phototransduction (Hardie 2001; Kumar 2001; Jeffery 2005; Friedrich et al. 2011). We also searched GO terms for vision-related terms such as: “eye”, “photoreceptor”, “phototransduction”, “R7”, “R8”, and “pigment”. Two-way ANOVAs were performed to examine the effects of sex, seasonal form, and sex × seasonal form using all 12 libraries for 203 contigs. To correct for multiple tests, FDR was calculated using the qvalue R package (Storey et al. 2003; Dabney 2014).

RT-PCR

12 animals (4 DSF, 4 WSF, 4 DSM, 4 WSM) were sacrificed 0-3 hours after eclosing. Eyes (retina + optic lobe) and brains (without optic lobe) were dissected, placed in RNAlater, and shipped to UC Irvine. Upon arrival, samples were placed in a freezer at -80°C. RNA was

extracted using TRIzol (Life Technologies, Grand Island, NY) and purified using the Nucleospin RNA II kit (Macherey-Nagel, Bethlehem, PA), which includes a DNase-treatment step. For RT-PCR, each 25 ml reaction had 2.5 µl Advantage 2 PCR buffer (Life Technologies, Grand Island, NY), 2.5 ml dNTPs (2 mM), 0.5 µl Choice-Taq Blue (Denville Scientific, South Plainfield, NJ), 0.5 ml (1:20 diluted) SuperScript® II Reverse Transcriptase (Life Technologies, Grand Island, NY), 1 µl 20x primer and probe mix (supplemental table S6, Integrated DNA Technologies, Coralville, IA), 17 µl H_2O and 1 µl RNA. The PCR reaction consisted of 45 cycles of (95°C for 30 sec, 55°C for 30 sec, and 68°C for 55 sec). We visualized amplification by running the PCR products on a 2% agrose gel.

ACKNOWLEDGMENTS

We thank Furong Yuan for RNA-Seq library preparation training and Andrew Everett for sharing his data for Table 1.1. We are grateful to Kevin Thornton, Ali Mortazavi, and Francisco Ayala for comments and advice on analyses. The Trinity assembly has been deposited in Dryad under data identifier doi:10.5061/dryad.f98s6, the raw sequencing reads have been deposited in ArrayExpress archive under accession E-MTAB-3887, and nucleotide sequences of the genes examined by RT-PCR have been deposited in GenBank under accession numbers KT781079-KT781084. This work was supported by the National Science Foundation (NSF) IOS-1145933 award and the Singapore Ministry of Education Award MOE2014-T2-1-146 to AM and by the NSF BEACON Award DBI-0939454 to ADB.

REFERENCES

Aspiras AC, Prasad R, Fong DW, Carlini DB, Angelini DR. 2012. Parallel reduction in expression of the eye development gene *hedgehog* in separately derived cave populations of the amphipod *Gammarus minus*. *J Evol Biol.* 25:995–1001.

Aubin-Horth N, Renn SCP. 2009. Genomic reaction norms: using integrative biology to understand molecular mechanisms of phenotypic plasticity. *Mol Ecol*. 18:3763–3780.

Baonza A, Casci T, Freeman M. 2001. A primary role for the epidermal growth factor receptor in ommatidial spacing in the *Drosophila* eye. *Curr Biol.* 11:396–404.

Battelle BA, Kempler KE, Parker AK, Gaddie CD. 2013. Opsin1-2, $G_q\alpha$ and arrestin levels at *Limulus* rhabdoms are controlled by diurnal light and a circadian clock. *J Exp Biol.* 216:1837–1849

Bear A, Monteiro A. 2013. Male courtship rate plasticity in the butterfly *Bicyclus anynana* is controlled by temperature experienced during the pupal and adult stages. *PLoS One* 8:e64061.

Bel Y, Jacobson KB, Silva FJ, Ferré J. 1992. Developmental and biochemical studies on the phenylalanine hydroxylation system in *Drosophila melanogaster*. *Insect Biochem Mol Biol.* 22:633–638.

Brakefield PM, Beldade P, Zwaan BJ. 2009. The African butterfly *Bicyclus anynana*: a model for evolutionary genetics and evolutionary developmental biology. *Cold Spring Harb. Protoc.* 4:1–9.

Brakefield PM, Reitsma N. 1991. Phenotypic plasticity, seasonal climate and the population biology of *Bicyclus* butterflies (Satyridae) in Malawi. *Ecol Entomol.* 16:291–303.

Briscoe AD, Bernard GD. 2005. Eyeshine and spectral tuning of long wavelength-sensitive rhodopsins: no evidence for red-sensitive photoreceptors among five Nymphalini butterfly species. *J Exp Biol.* 208:687–696.

Briscoe AD, Chittka L. 2001. The evolution of color vision in insects. *Annu Rev Entomol.* 46:471–510.

Briscoe AD. 2008. Reconstructing the ancestral butterfly eye: focus on the opsins. *J Exp Biol.* 211:1805–1813.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Chang JL, Lin HV, Blauwkamp TA, Cadigan KM. 2008. Spenito and Split ends act redundantly to promote Wingless signaling. *Dev Biol.* 314:100–111.

Chen R, Amoui M, Zhang Z, Mardon G. 1997. Dachshund and eyes absent proteins form a complex and function synergistically to induce ectopic eye development in *Drosophila*. *Cell* 91:893–903.

CLC bio. 2012. CLC Genomics Workbench. www.clcbio.com

Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, Barluenga M, Salzburger W. 2013. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Mol Ecol.* 22:670–684.

Crispo E. 2007. The Baldwin Effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution* 61:2469–2479.

Culver DC, Pipan T. 2009. The biology of caves and other subterranean habitats. New York: Oxford University Press.

Dabney A, Storey JD. 2013. qvalue: Q-value estimation for false discovery rate control. R package version 1.36.0.

Das Gupta M, Chan SK, Monteiro A. 2015. Natural loss of *eyeless/Pax6* in eyes of *Bicyclus anynana* adult butterflies likely leads to exponential decrease of eye fluorescence in transgenics. PLoS One 10(7):e0123882.

Dauwalder B, Tsujimoto S, Moss J, Mattox W. 2002. The *Drosophila takeout* gene is regulated by the somatic sex-determination pathway and affects male courtship behavior. *Genes Dev*. 16:2879–2892.

Doroquez DB, Orr-Weaver TL, Rebay I. 2007. Split ends antagonizes the Notch and potentiates the EGFR signaling pathways during *Drosophila* eye development. *Mech Dev.* 124:792–806.

Eisen MB, Spellman PT, Brown PO, Botsein D.1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci*. *USA* 95:14863–14868.

Everett A, Tong X, Briscoe AD, Monteiro A. 2012. Phenotypic plasticity in opsin expression in a butterfly compound eye complements sex role reversal. *BMC Evol Biol.* 12:232.

Fitzpatrick BM. 2012. Underappreciated consequences of phenotypic plasticity for ecological speciation. *Int J Ecol*. 2012:256017.

Flaherty MS, Zavadil J, Ekas LA, Bach EA. 2009. Genome-wide expression profiling in the *Drosophila* eye reveals unexpected repression of notch signaling by the JAK/STAT pathway. *Dev Dynam*. 238:2235–2253.

Frankfort BJ, Mardon G. 2004. Senseless represses nuclear transduction of Egfr pathway activation. *Development* 131:563–570.

Freeman M. 1997. Cell determination strategies in the *Drosophila* eye. *Development* 124:261–270.

Friedrich M, Chen R, Daines B, Bao R, Caravas J, Rai PK, Zagmajster M, Peck SB. 2011. Phototransduction and clock gene expression in the troglobiont beetle *Ptomaphagus hirtus* of Mammoth cave. *J Exp Biol.* 214:3532–3541.

Friedrich M. 2003. Evolution of insect eye development: first insights from fruit fly, grasshopper and flour beetle. *Integr Comp Biol.* 43:508–521.

Friedrich M. 2013. Biological clocks and visual systems in cave-adapted animals at the dawn of speleogenomics. *Integr Comp Biol.* 53:50–67.

Fong DW, Kane TC, Culver DC. 1995. Vestigialization and causes of vestigialization. *Annu Rev Ecol Syst.* 26:249–268.

Ghalambor CK, McKay JK, Carroll SP, Renick DN. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct Ecol*. 21:394–407.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 29:644–652.

Hardie RC. 2001. Phototransduction in *Drosophila melanogaster*. *J Exp Biol.* 204:3403–3409.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 8:1494-1512.

Higashijima S, Kojima T, Michiue T, Ishimaru S, Emori Y, Saigo K. 1992. Dual *Bar* homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes Dev.* 6:50–60.

Hofmann CM, O'Quin KE, Smith AR, Carleton KL. 2010. Plasticity of opsin gene expression in cichlids from Lake Malawi. *Mol Ecol*. 19:2064–2074.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc.* 4:44–57.

Jeffery WR. 2005. Adaptive evolution of eye degeneration in the Mexican blind cavefish. *J Hered*. 96:185–196.

Kametaka S, Kametaka A, Yonekura S, Haruta M, Takenoshita S, Goto S, Waguri S. 2012. AP-1 clathrin adaptor and CG8538/Aftiphilin are involved in Notch signaling during eye development in *Drosophila melanogaster*. *J Cell Sci*. 125:634–648.

Koch PB, Brakefield PM, Kesbeke F. 1996. Ecdysteroids control eyespot size and wing color pattern in the polyphenic butterfly *Bicyclus anynana* (Lepidoptera: Satyridae). *J Insect Physiol*. 42:223–230.

Kokubo H, Ueno K, Amanai K, Suzuki Y. 1997. Involvement of the *Bombyx Scr* gene in development of the embryonic silk gland. *Dev. Biol.* 186:46-57

Kooi RE, Brakefield PM. 1999. The critical period for wing pattern induction in the polyphenic tropical butterfly *Bicyclus anynana* (Satyrinae). *J Insect Physiol*. 45:201–212.

Kumar JP. 2001. Signalling pathways in *Drosophila* and vertebrate retinal development. *Nat Rev Genet*. 2:846–857.

Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA. 2009. Relaxed selection in the wild. *Trends Ecol Evol.* 24:487–496.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.

Marygold SJ, Leyland PC, Seal RL, Goodman JL, Thurmond JR, Strelets VB, Wilson RJ, The FlyBase Consortium. 2013. FlyBase: improvements to the bibliography. *Nucl Acids Res.* 41:D751–D757.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucl Acids Res.* 40:4288–4297.

Meng F, Braasch I, Phillips JB, Lin X, Titus T, Zhang C, Postlethwait JH. 2013. Evolution of the eye transcriptome under constant darkness in *Sinocyclocheilus* cavefish. *Mol Biol Evol.* 30:1527–1543.

Mikeladze-Dvali T, Wernet MF, Pistillo D, Mazzoni EO, Teleman AA, Chen Y, Cohen S, Desplan C. 2005. The growth regulators *warts*/*lats* and melted interact in a bistable loop to specify opposite fates in *Drosophila* R8 photoreceptors. *Cell* 122:775–87.

Monteiro A, Tong X, Bear A, Liew SF, Bhardwaj S, Wasik BR, Dinwiddie A, Bastianelli C, Cheong WF, Wenk MR, Cao H, Prudic KL. In press. Differential expression of ecdysone receptor leads to variation in phenotypic plasticity across serial homologs PLoS Genetics.

Moran NA. 1992. The evolutionary maintenance of alternative phenotypes. *Am Nat*. 139:971–989.

Mosley-Bishop KL, Li Q, Patterson K, Fischer JA. 1999. Molecular analysis of the *klarsicht* gene and its role in nuclear migration within differentiating cells of the *Drosophila* eye. *Curr Biol*. 9:1211–1220.

Nijhout FH. 2003. Development and evolution of adaptive polyphenisms. *Evol Dev*. 5:9–18.

Niven JE, Anderson JC, Laughlin SB. 2007. Fly photoreceptors demonstrate energy-information trade-offs in neural coding. *PLoS Biol.* 5:e116.

Niven JE, Laughlin SB. 2008. Energy limitation as a selective pressure on the evolution of sensory systems. *J Exp Biol.* 211:1792–1804.

Niven JE. 2014. Neural energetics: Hungry flies turn down the visual gain. *Curr Biol*. 24:R313–R315.

Pappu KS, Chen R, Middlebrooks BW, Woo C, Heberlein U, Mardon G. 2003. Mechanism of *hedgehog* signaling during *Drosophila* eye development. *Development* 130:3053–3062.

Passalacqua KD, Hrycaj S, Mahfooz N, Popadic A. 2010. Evolving expression patterns of the homeotic gene *Scr* in insects. *Int J Dev Biol*. 54:897–904.

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP. 2010. Phenotypic plasticity's impacts on diversification and speciation. *Trends Ecol Evol*. 25:459–467.

Pigliucci M, Murren CJ, Schlichting CD. 2006. Phenotypic plasticity and evolution by genetic assimilation. *J Exp Biol*. 209:2362–2367.

Ploner, A. 2012. Heatplus: Heatmaps with row and/or column covariates and colored clusters. Version 2.6.0.

Price TD, Qvarnström A, Irwin DE. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proc R Soc B Biol Sci*. 270:1433–1440.

Prudic KL, Jeon C, Cao H, Monteiro A. 2011. Developmental plasticity in sexual roles of butterfly species drives mutual sexual ornamentation. *Science* 331:73–75.

R Core Team. 2013. A Language and Environment for Statistical Computing. Vienna, Austria. http://www.R-project.org

Rivera AS, Pankey MS, Plachetzki DC, Villacorta C, Syme AE, Serb JM, Omilian AR, Oakley TH. 2010. Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. *BMC Evol Biol*. 10:123.

Robertson KA, Monteiro A. 2005. Female *Bicyclus anynana* butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. *Proc R Soc B Biol Sci*. 272:1541–1546.

Robinson MD, McCarthy DJ and Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-Seq data. *Genome Biol*. 11:R25.

Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23:2881-2887.

Robinson MD, Smyth GK. 2008. Small sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9:321-332.

Rogers BT, Peterson MD, Kaufman TC. 1997. Evolution of the insect body plan as revealed by Sex combs reduced expression pattern. *Development* 124:149–157

Sánchez L, Guerrero I. 2001. The development of the *Drosophila* genital disc. *BioEssays* 23:698–707.

Sasagawa H, Narita R, Kitagawa Y, Kadowaki T. 2003. The expression of genes encoding visual components is regulated by a circadian clock, light environment and age in the honeybee (*Apis mellifera*). *Eur J Neurosci*. 17: 963–970.

Schlichting CD, Smith H. 2002. Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. *Evol Ecol*. 16:189–211.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikoski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13:2498–504.

Smith G, Briscoe AD. 2015. Molecular evolution and expression of the CRAL_TRIO protein family in insects. *Insect Biochem Molec Biol.* 62:168-173.

Smith G, Fang Y, Liu X, Kenny J, Cossins AR, de Oliveira CC, Etges WJ, Ritchie MG. 2013. Transcriptome-wide expression variation associated with environmental plasticity and mating success in cactophilic *Drosophila mojavensis*. *Evolution* 67:1950–1963.

Spaethe J, Briscoe AD. 2005. Molecular characterization and expression of the UV opsin in bumblebees: three ommatidial subtypes in the retina and a new photoreceptor organ in the lamina. *J Exp Biol.* 208: 2347–2361.

Steinger T, Roy BA, Stanton ML. 2003. Evolution in stressful environments II: adaptive value and costs of plasticity in response to low light in *Sinapis arvensis*. *J Evol Biol*. 16:313–323.

Storey JD, Tibshirani R 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci*. *USA* 100: 9440-9445.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 28: 2731–2739.

Tan S, Amos W, Laughlin SB. 2005. Captivity selects for smaller eyes. *Curr Biol*. 15:R540–2.

Tanaka K, Barmina O, Sanders LE, Arbeitman MN, Kopp A. 2011. Evolution of sex-specific traits through changes in HOX-dependent *doublesex* expression. *PLoS Biol.* 9: e1001131.

Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in RNA-Seq: a matter of depth. *Genome Res*. 21:2213–2223.

The UniProt Consortium. 2013. Activities at the Universal Protein Resource (UniProt). *Nucl Acids Res*. 42:D191-D198.

Tsai Y-C, Sun YH. 2004. Long-range effect of upd, a ligand for Jak/STAT pathway, on cell cycle in *Drosophila* eye development. *Genesis* 39:141–153.

Wang T, Montell C. 2005. Rhodopsin formation in *Drosophila* is dependent on the PINTA retinoid-binding protein. *J Neurosci*. 25:5187–5194.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 10:57–63.

West-Eberhard M. 2003. Developmental Plasticity and Evolution. New York: Oxford University Press.

Yamamoto Y, Stock DW, Jeffery WR. 2004. Hedghog signaling controls eye degeneration in blind cavefish. *Nature* 431:844–874.

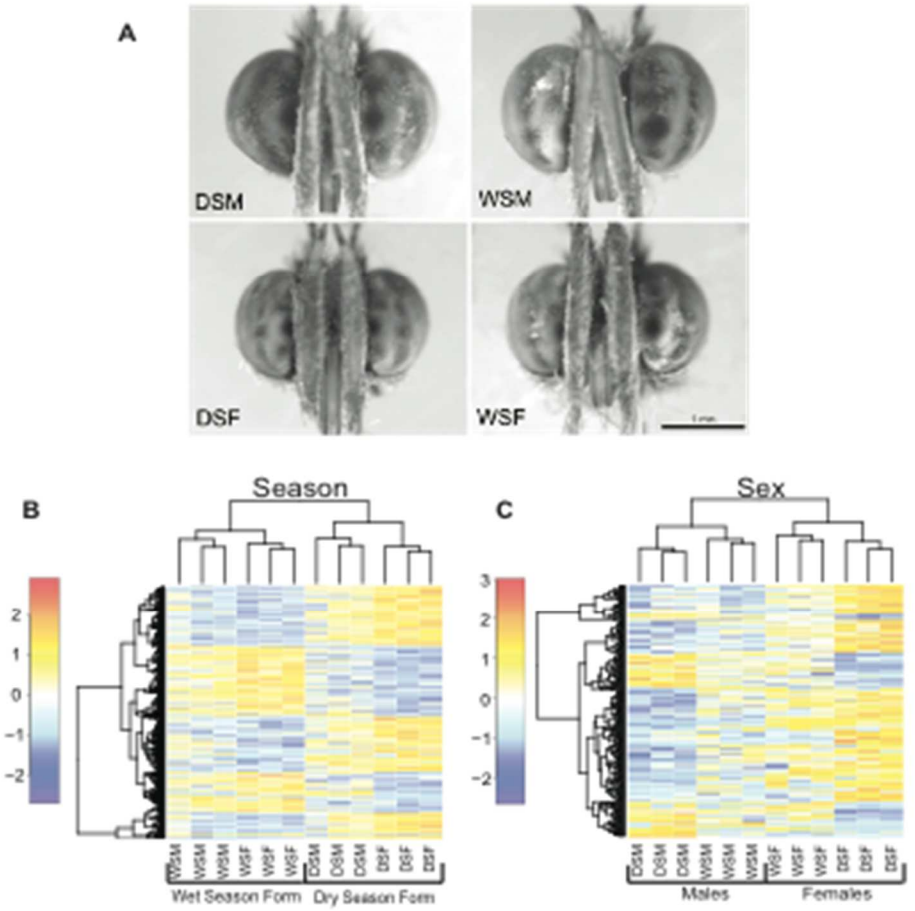


Figure 1.1. Photographs of adult *Bicyclus anynana* heads and heatmaps of diffentially expressed (DE) contigs by seasonal form and sex. (*a*) Eye size varies between wet season (WS) and dry season (DS) forms and between males (M) and females (F). Heatmaps of DE contigs by seasonal form (*b*) and sex (*c*). Color bar indicates scaled logCPM (log counts per million).

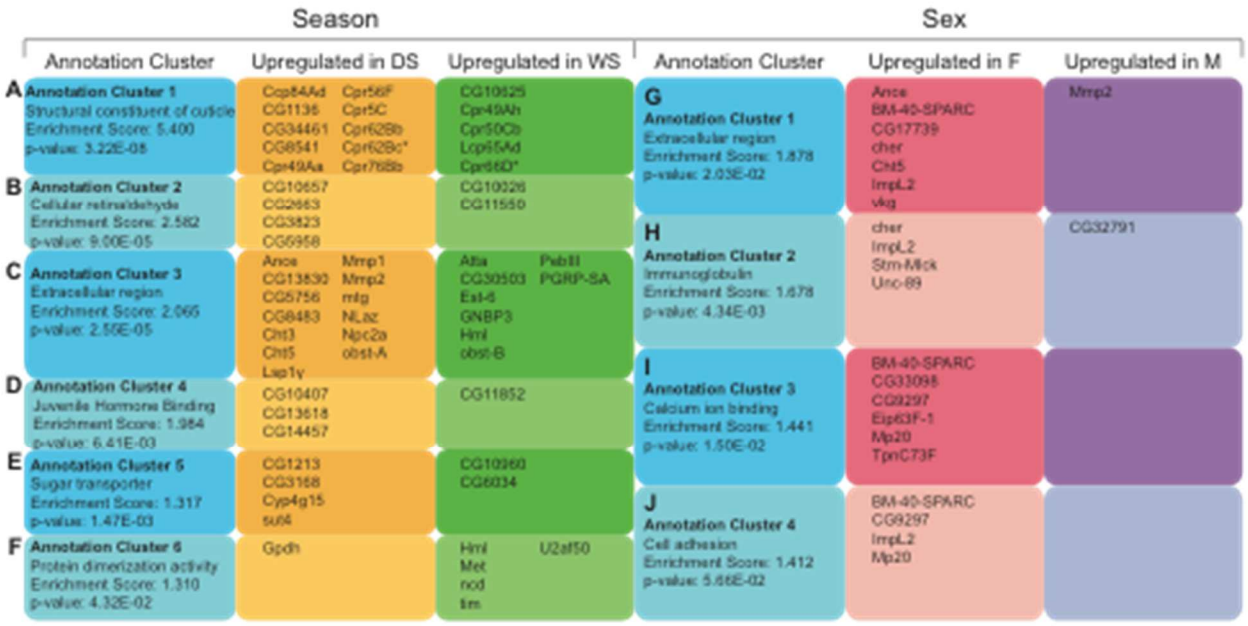


Figure 1.2. Functional enrichment of differentially expressed genes. A majority of contigs were upregulated in DS forms and in females. Enrichment clusters of DE genes by season included contigs homologous to: (*a*) insect cuticle proteins, (*b*) cellular retinaldehyde binding and alpha-tocopherol transport-like proteins, (*c*) extracellular region, aminoglycan and chitin metabolism proteins, (*d*) odorant and juvenile hormone binding proteins, (*e*) sugar transport proteins and (*f*) proteins with dimerization activity and binding. Enrichment clusters of DE genes by sex included homologs to: (*g*) extracellular region, (*h*) immunoglobulin, (*i*) calcium ion binding and (*j*) cell adhesion. Asterisks denote contigs that were up-regulated in DSF and WSM. Gene names are based on *Drosophila* homologs.

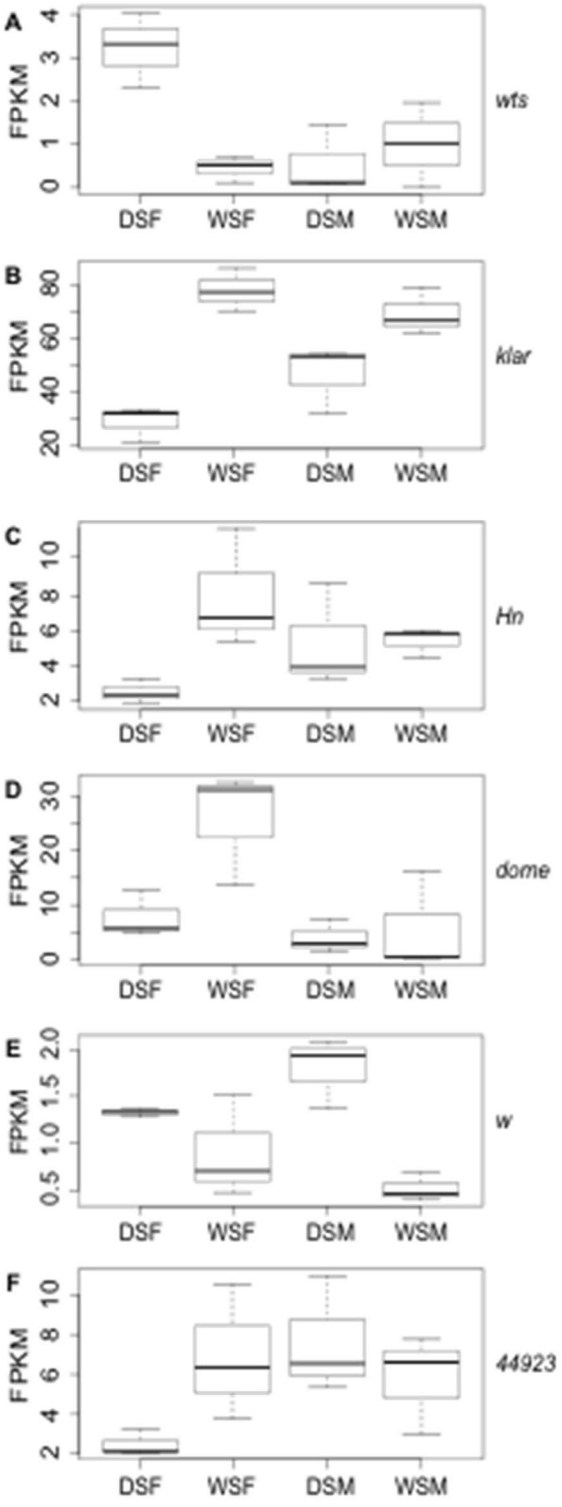

A
FPKM
0 1 2 3 4
DSF WSF DSM WSM
wfs
B
FPKM
20 40 60 80
DSF WSF DSM WSM
klar
C
FPKM
2 4 6 8 10
DSF WSF DSM WSM
Hn
D
FPKM
0 10 20 30
DSF WSF DSM WSM
dome
E
FPKM
0.5 1.0 1.5 2.0
DSF WSF DSM WSM
w
F
FPKM
2 4 6 8 10
DSF WSF DSM WSM
44923

Figure 1.3. Expression of differentially expressed genes across seasonal forms. (*a*) *Warts* (*wts*) has increased expression in DSF relative to other forms. (*b*) *Klarsicht* (*klar*) expression is decreased in DSF relative to WSF. (*c*) *Henna* (*Hn*) expression is decreased in DSF relative to WSF. (*d*) *Domeless* (*dome*) expression is also decreased in DSF relative to WSF. (*e*) *comp44923* has decreased expression in DSF relative to other forms. (*f) White* (*w*) expression is increased in DSM relative to WSM.

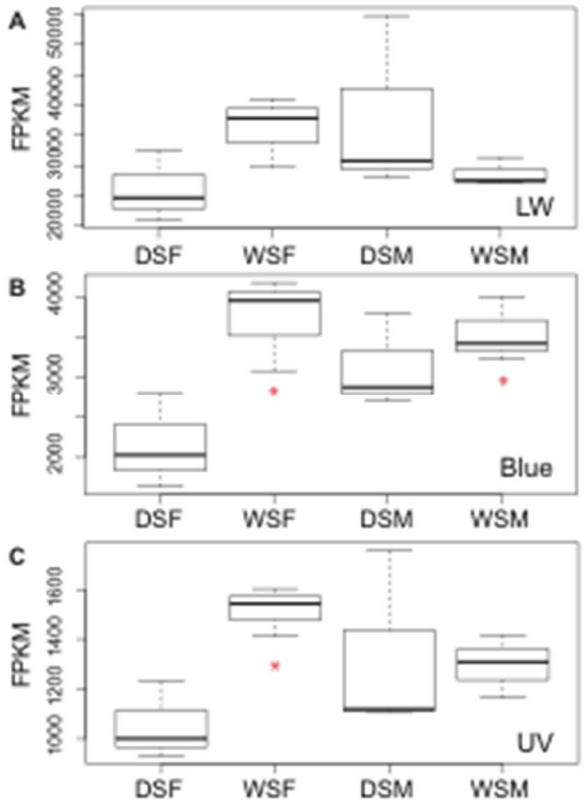


Figure 1.4. Opsin expression in *B. anynana* seasonal forms. Box plots show FPKM values. Thick lines represent medians. Whiskers represent maximum and minimum values. Asterisks denote differential expression relative to DSF using one-way ANOVAs (*a*) Long-wavelength (*LWRh)* opsin expression. (*b*) Blue (*BRh*) opsin expression. (*c*) Ultraviolet (*UVRh*) opsin expression.

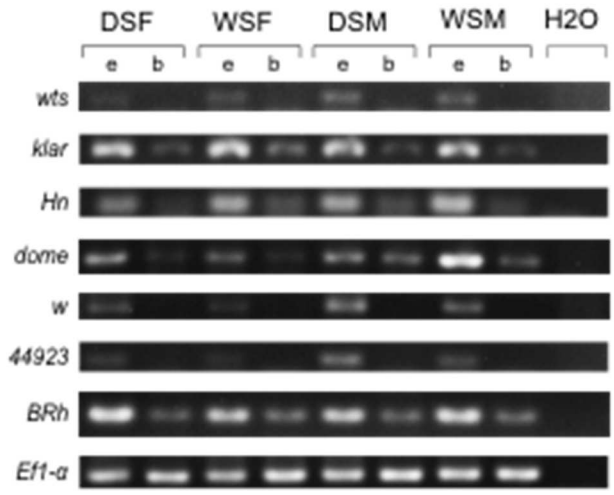


Figure 1.5. Reverse-transcription PCR (RT-PCR) of eye and brain tissue. RT-PCR in dry season female (DSF), wet season female (WSF), dry season male (DSM), and wet season male (WSM) eye (e) and brain (b) tissue show that *warts, comp44923* and *white* are only expressed in eye tissue. *Klarsicht* (*klar*), *henna* (*hn*), and *domeless* (*dome*) are expressed in both eye and brain tissue. *Blue rhodopsin* (*BRh*) and *elongation factor 1-α* (*EF1-α*) are positive controls and are both found in eye and brain tissue.

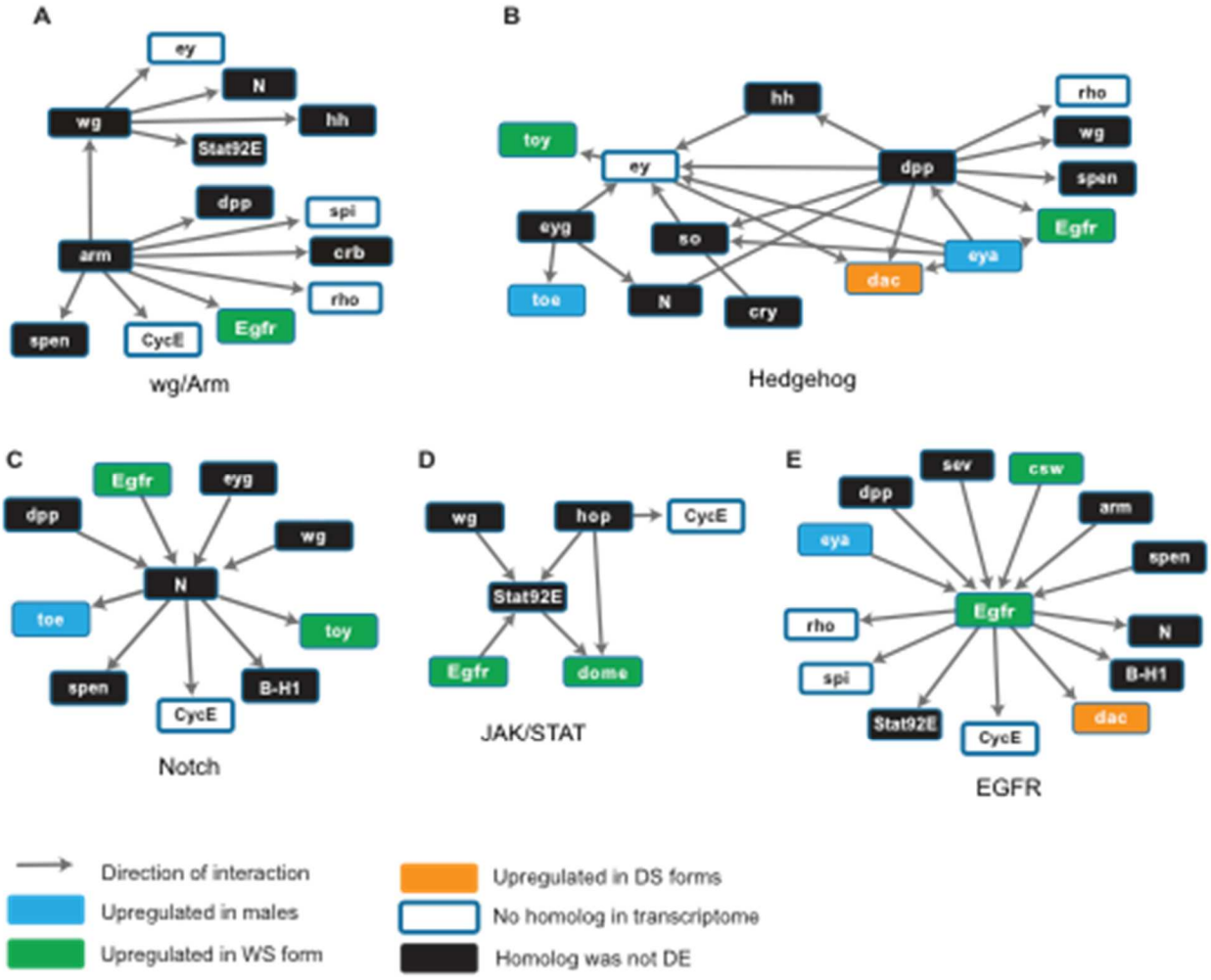


Figure 1.6. Eye development gene networks. (*a*) Wingless/Armadillo (wg/Arm) signaling pathway. (*b*) Hedgehog (hh) signaling pathway. (*c*) Notch (N) signaling pathway. (*d*) Janus Kinase/Signal Transducer and Activator of Transcription (JAK/STAT) signaling pathway. (*e*) Epidermal growth factor receptor (Egfr) signaling pathway. Arrows reflect direction of interaction from Flybase data for Cytoscape (Shannon et al. 2003). Blue nodes represent genes up-regulated in males. Orange nodes represent genes up-regulated in DS forms. Green nodes represent genes up-regulated in WS forms. Black nodes represent genes with a homolog to the *Drosophila* gene in our transcriptome but which are not differentially expressed (DE) and white nodes indicate genes where no homolog was found in our assembly.

Table 1.1. Summary of findings from Everett et al. (2012).

Season	Sex	Behavior	Mean Eye Size	Mean Facet Number	Mean Facet Size	Relative *BRh* expression	Relative *UVRh* expression	Relative *LWRh* expression
Dry	F	non-choosy	0.87± 0.06 mm²	2311 ± 326	348.8 ± 21 µm²	-0.80†‡	0.78‡	0.01†‡
Dry	M	choosy	1.09 ± 0.065 mm²	2727 ± 242.5	396.8 ± 13 µm²	0.61	2.01	0.76
Wet	F	choosy	0.98 ± 0.06 mm²	2857 ± 326	342.9 ± 21 µm²	0.83	2.39	0.38
Wet	M	non-choosy	1.33 ± 0.085 mm²	3541 ± 352.5	362.8 ± 19 µm²	0.53	2.12	0.44
Global gene prediction			Eye developmental genes and eye differentiation genes will be up-regulated or down-regulated in males relative to females and WS forms relative to DS forms			Additional phototransduction genes will be down-regulated in dry season females.		

Note – Opsin expression was quantified using qPCR and by normalizing to 18S rRNA then against the normalized opsin levels of a randomly picked sample using $2^{-\Delta\Delta CT}$ method (see Everett et al. 2012 for details). N=3 biological and n=2 technical replicates were performed.

† $p < 0.05$ for DSF vs. DSM
‡ $p < 0.05$ for DSF vs. WSF

Table 1.2. Summary of the total number of differentially expressed (DE) contigs and unique gene ontology (GO) terms discovered in analyses.

Libraries used	Contrasts	DE contigs	GO terms
All 12	Sex	290	77
All 12	Season	722	229
All 12	Interaction	111	23
Females only	Season	790	267

Table 1.3. BLAST results for contigs with opsin-like expression patterns.

Contig ID	Description	Top Hit
comp33544_c0	Centromere protein 1 and 1-like	[*Danaus plexippus*] hypothetical protein KGM_06860
comp33612_c0	Sugar transporter	[*Bombyx mori*] sugar transporter 4
comp42445_c0	FAM50 homolog; neurogenesis	[*Danaus plexippus*] hypothetical protein KGM_09648
comp42682_c0	Pigment binding; small molecule binding	[*Danaus plexippus*] Bombyrin
comp43633_c0	Chemosensory binding; odorant binding; serine/threonine kinase	[*Bombyx mori*] uncharacterized protein LOC101743765
comp44249_c0	Serine protease inhibitor; antitrypsin isoform; peptidase activity	[*Manduca sexta*] serpin 1
comp44923_c0	Transporter activity; alpha-tocopherol transfer	[*Danaus plexippus*] putative CRAL/TRIO domain-containing protein
comp46683_c0	Secreted protein; salivary cys-rich secreted protein	[*Danaus plexippus*] hypothetical protein KGM_05173
comp47781_c0	Heparin sulfate O-sulfotransferase-like	[*Bombyx mori*] heparin sulfate O-sulfotransferase-like
comp48361_c0	BTB/POZ domain-containing protein	[*Bombyx mori*] BTB/POZ domain-containing protein 2-like
comp48541_c0	Nicotinamide riboside kinase	[*Bombyx mori*] nicotinamide riboside kinase 2-like isoform X1
comp49621_c0	L-asparaginase and like; lysophospholipase	[*Danaus plexippus*] lyso
comp49695_c0	Glucose dehydrogenase; glucose dehydrogenase precursor and acceptor	[*Danaus plexippus*] hypothetical protein KGM_15606
comp50819_c1	Hematological and neurological expressed 1-like protein	[*Danaus plexippus*] hypothetical protein KGM_13882
comp52506_c0	Inter-alpha-trypsin inhibitor heavy chain H4	[*Bombyx mori*] inter-alpha-trypsin inhibitor heavy chain H4-like
comp115040_c0	Phenylalanine hydroxylase; protein henna-like	[*Danaus plexippus*] phenylalanine hydroxylase

Table 1.4. Significant *P* values of two-factor ANOVAs for eye development, phototransduction, and eye pigment genes.

	Gene	Sex	Season	Interaction
Eye Development	*a*	0.358	**0.017**	0.418
	Afti	0.546	**0.003**	0.549
	aop	**0.032**	**0.014**	**0.017**
	AP-1sigma	**0.001**	**0.035**	0.527
	bab2	0.840	**0.001**	0.888
	boi	0.655	**0.036**	0.306
	Bx42	0.289	**0.013**	0.195
	csw	0.483	**0.006**	0.800
	dac	0.457	**0.004**	0.401
	Dad	0.317	**0.020**	0.538
	Doa	0.540	**0.020**	0.890
	E(spl)mbeta-HLH	0.170	**0.017**	0.510
	E(spl)mgamma-HLH	**0.025**	**0.010**	0.230
	Egfr	0.268	**0.044**	0.098
	eya	**0.040**	0.062	0.900
	gl	**0.036**	**0.011**	0.134
	Gp150	**0.006**	**0.031**	0.920
	holn1	**0.015**	0.919	0.720
	hth	**0.005**	0.137	0.916
	kay	**0.001**	**0.015**	0.110
	klar	0.423	**0.000**	**0.040**
	lin-52	0.076	**0.012**	0.250
	PDZ-GEF	0.570	**0.015**	0.380
	peb	0.114	**0.014**	0.119
	pelo	0.258	**0.013**	0.320
	rst	**0.011**	0.646	0.976
	sca	0.378	**0.012**	0.312
	scrib	0.179	0.081	**0.043**
	skd	0.070	**0.005**	0.667
	ssh	**0.049**	**0.032**	0.090
	Tak1	0.830	0.188	**0.007**
	tio	**0.045**	**0.004**	**0.035**
	toe	**0.008**	0.750	0.826
	toy	0.709	**0.011**	0.828
Phototransduction	*Arr2*	0.422	0.250	**0.039**
	CG11426	0.931	0.385	**0.009**

	cry	**0.011**	0.389	0.975
	Pld	0.134	**0.045**	0.112
	rdgA	**0.030**	0.857	0.876
	rdgC	0.667	**0.038**	0.422
	BRh	0.246	**0.013**	0.102
	shakB	**0.009**	**0.001**	**0.047**
Eye Pigment	*dor*	0.066	**0.003**	0.387
	Dysb	0.657	**0.012**	0.761
	lt	0.960	**0.014**	0.057
	or	0.131	**0.023**	0.680
	p	**0.009**	0.191	0.553
	Pu	0.642	0.652	**0.036**
	st	**0.002**	0.337	0.219
	w	**0.022**	**0.010**	0.560

Note -- Bold indicated *P* values < 0.05. However, these values were not significant after FDR correction.

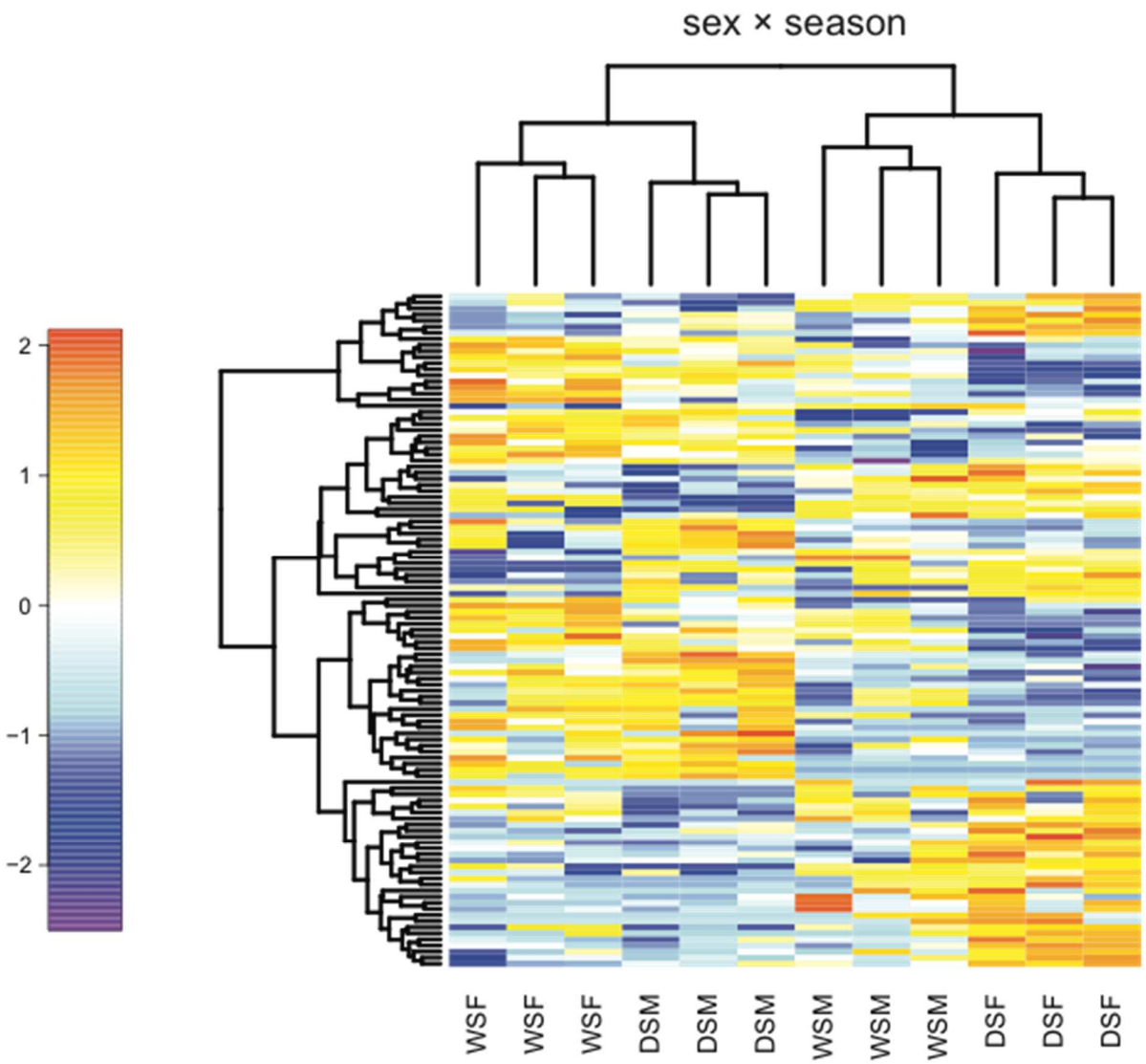


Figure S1.1. Heatmap of differentially expressed contigs showing a sex by seasonal form interaction. Differentially expressed (DE) contigs for sex×seasonal form interaction are the result of a two factor generalized linear model using 12 libraries. Color indicates logCPM between treatment types.

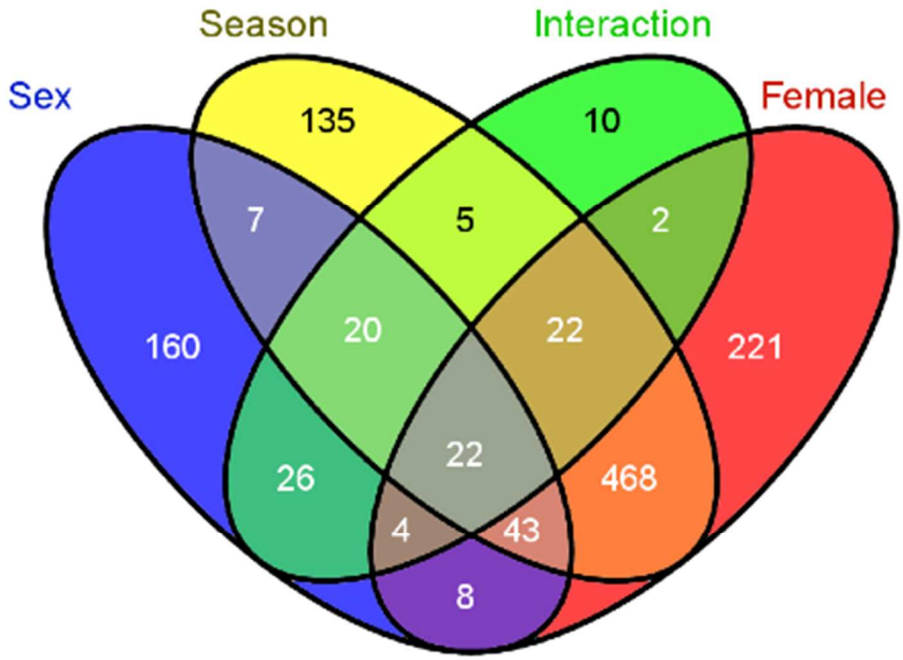


Figure S1.2. Venn diagram comparing two factor and single factor comparison. Venn diagram shows number of unique and overlapping DE contigs by sex (blue), season (yellow), and interaction (green) for a two factor generalized linear model using 12 libraries. We also included DE contigs for a single factor comparison between female seasonal forms.

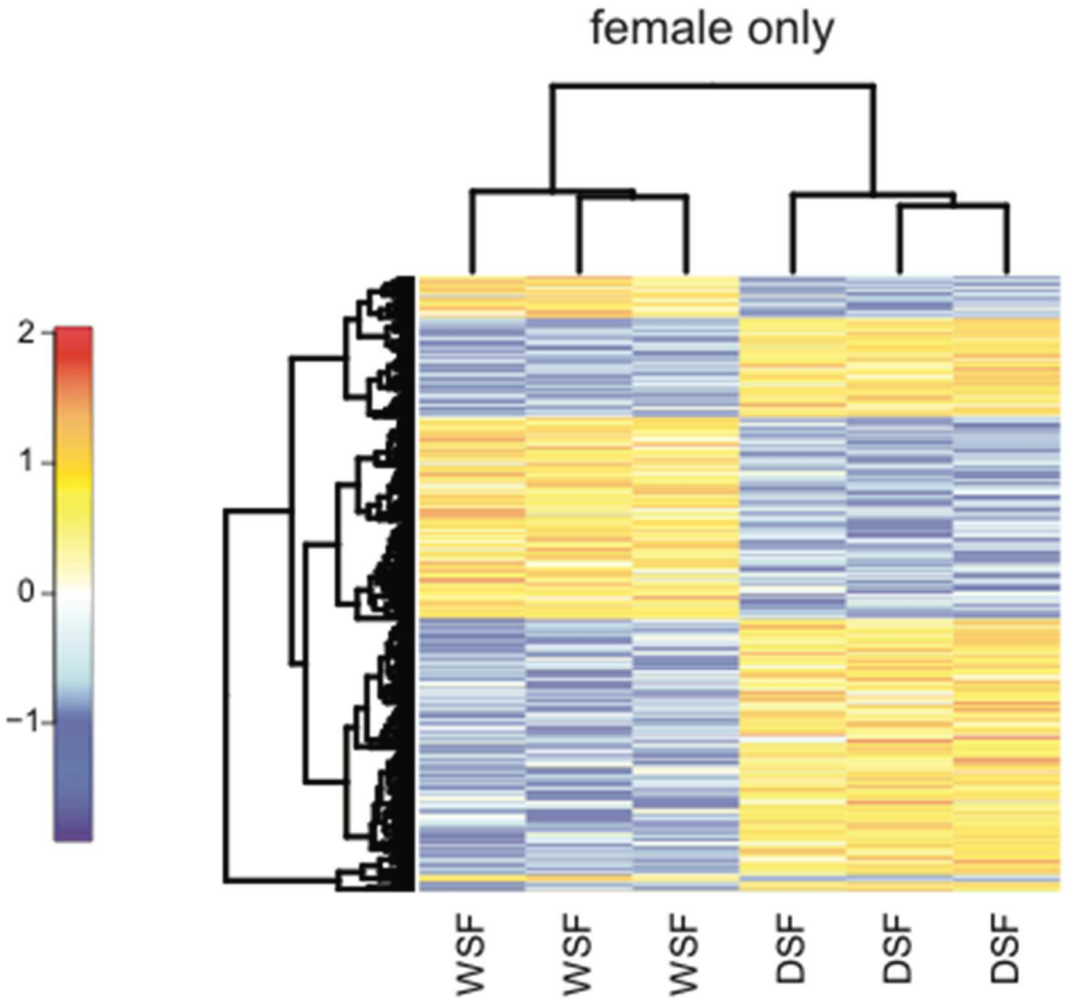


Figure S1.3. Heatmap of differentially expressed contigs between female seasonal forms. Heatmap of 790 DE contigs from single factor analysis of female seasonal forms (FDR < 0.05), approximately two-thirds of these contigs are up-regulated in dry season females.

Table S1.1. Library mapping statistics.

Treatment Type	Library ID	Rearing Temperature	100 bp PE Reads	Reads Processed	% align	Reads Align
Dry Season Female (DSF)	BA21_DS_F	17°C	18,643,644	16,347,019	85.33	13,948,765
	BA25_DS_F	17°C	15,622,467	13,825,745	83.15	11,495,948
	BA29_DS_F	17°C	14,283,898	12,616,223	83.41	10,523,190
Dry Season Male (DSM)	BA23_DS_M	17°C	16,259,972	14,318,037	84.98	12,024,429
	BA27_DS_M	17°C	15,696,976	13,810,803	84.39	11,655,214
	BA31_DS_M	17°C	17,282,610	15,256,682	83.67	12,764,868
Wet Season Female (WSF)	BA22_WS_F	27°C	15,669,579	13,836,887	85.23	11,793,555
	BA26_WS_F	27°C	16,060,400	14,153,248	83.68	11,843,165
	BA30_WS_F	27°C	15,766,658	13,915,314	84.54	11,764,412
Wet Season Male (WSM)	BA24_WS_M	27°C	15,928,027	14,067,886	84.98	11,954,219
	BA28_WS_M	27°C	15,746,421	13,903,314	83.37	11,591,823
	BA32_WS_M	27°C	16,294,576	14,073,211	83.07	11,690,586

Table S1.2. BLAST results for Trinity assembly peptide sequences to *Drosophila* translated sequences.

https://academic.oup.com/mbe/article/33/1/79/2579302#supplementary-data

Table S1.3. Functional enrichment for DE contigs.

Annotation Clusters	Term	Count	P-value	Genes	FDR
Annotation Cluster A	structural constituent of cuticle	15	3.22E-08	CG34461, CPR76BB, CCP84AD, CPR5C, CPR66D, CG10625, CPR49AH, CG8541, CPR62BC, CPR49AA, CPR62BB, CPR56F, CPR50CB, CG1136, LCP65AD	9.13E-06
Enrichment Score: 5.400	Insect cuticle protein	14	3.74E-08	CG34461, CPR76BB, CCP84AD, CPR5C, CPR66D, CG10625, CPR49AH, CPR62BC, CPR49AA, CPR62BB, CPR56F, CPR50CB, CG1136, LCP65AD	1.32E-05
	structural constituent of chitin-based cuticle	13	6.81E-07	CG34461, CPR76BB, CCP84AD, CPR5C, CPR66D, CPR49AH, CG8541, CPR49AA, CPR62BC, CPR62BB, CPR56F, CPR50CB, LCP65AD	1.93E-04
Annotation Cluster B	SEC14	6	9.00E-05	CG5958, CG10026, CG2663, CG10657, CG3823, CG11550	0.007
Enrichment Score: 2.582	Cellular retinaldehyde binding/alpha-tocopherol transport	5	2.32E-04	CG5958, CG10026, CG2663, CG10657, CG3823	0.079
	Cellular retinaldehyde-binding/triple function, C-terminal	6	2.51E-04	CG5958, CG10026, CG2663, CG10657, CG3823, CG11550	0.085
Annotation Cluster C	extracellular region	21	2.55E-05	PEBIII, CHT5, CHT3, HML, OBST-B, OBST-A, ATTA, LSP1GAMMA, MMP2, MMP1, CG13830, NLAZ, EST-6, CG5756, CG8483, NPC2A, CG30503, GNBP3, PGRP-SA, MTG, ANCE	0.003
Enrichment Score: 2.065	ChtBD2	7	1.86E-03	CG5756, CHT5, CHT3, HML, OBST-B, OBST-A, MTG	0.129
	aminoglycan metabolic process	8	2.15E-03	CG5756, CHT5, CHT3, HML, OBST-B, OBST-A, MTG, PGRP-SA	0.826
Annotation Cluster D	JHBP	4	6.41E-03	CG14457, CG10407, CG13618, CG11852	0.379
Enrichment Score: 1.984	Odorant binding protein	4	1.14E-02	CG14457, CG10407, CG13618, CG11852	0.983
	Hormone binding	4	1.53E-02	CG14457, CG10407, CG13618, CG11852	0.996
Annotation Cluster E	Sugar transporter, conserved site	6	1.47E-03	CG10960, CYP4G15, CG3168, CG1213, SUT4, CG6034	0.405

Enrichment Score: 1.317	General substrate transporter	4	1.97E-02	CG42269, CG10960, CG1213, SUT4	0.999
	Sugar/inositol transporter	3	2.43E-02	CG10960, CG1213, SUT4	1.000
Annotation Cluster F Enrichment Score: 1.310	protein dimerization activity	6	4.32E-02	TIM, NCD, HML, MET, U2AF50, GPDH	1.000
	identical protein binding	7	4.99E-02	NCD, HML, MET, GPDH, L(2)EFL, PYD3, HGO	1.000
	protein homodimerization activity	4	5.44E-02	NCD, HML, MET, GPDH	1.000
Annotation Cluster G Enrichment Score: 1.878	extracellular region part	6	1.35E-03	IMPL2, VKG, CHER, MMP2, ANCE, BM-40-SPARC	0.107
	extracellular region	8	2.03E-02	CHT5, IMPL2, CG17739, VKG, CHER, MMP2, ANCE, BM-40-SPARC	0.822
	proteinaceous extracellular matrix	3	3.18E-02	VKG, MMP2, BM-40-SPARC	0.934
Annotation Cluster H Enrichment Score: 1.678	Immunoglobulin I-set	4	4.11E-03	IMPL2, UNC-89, STRN-MLCK, CG32791	0.412
	Immunoglobulin-like fold	5	4.34E-03	IMPL2, UNC-89, STRN-MLCK, CG32791, CHER	0.429
	IGc2	4	1.49E-02	IMPL2, UNC-89, STRN-MLCK, CG32791	0.344
Annotation Cluster I Enrichment Score: 1.441	calcium ion binding	6	1.50E-02	CG9297, MP20, EIP63F-1, CG33098, TPNC73F, BM-40-SPARC	0.869
	EF-Hand type	4	1.96E-02	EIP63F-1, CG33098, TPNC73F, BM-40-SPARC	0.923
	EF-HAND 1	4	2.34E-02	EIP63F-1, CG33098, TPNC73F, BM-40-SPARC	0.953
Annotation Cluster J Enrichment Score: 1.412	calcium ion binding	6	1.50E-02	CG9297, MP20, EIP63F-1, CG33098, TPNC73F, BM-40-SPARC	0.869
	cell adhesion	4	5.66E-02	IMPL2, CG9297, MP20, BM-40-SPARC	1.000
	biological adhesion	4	6.82E-02	IMPL2, CG9297, MP20, BM-40-SPARC	1.000
Annotation Cluster K Enrichment Score: 1.245	endopeptidase activity	4	2.84E-02	CG4998, CG9737, NEP2, MMP2	0.750
	hydrolase	6	5.56E-02	CG4998, BTV, CG9737, NEP2, CG11438, MMP2	0.663
	peptidase activity, acting on L-amino acid peptides	4	5.98E-02	CG4998, CG9737, NEP2, MMP2	0.948
Annotation Cluster L	Insect cuticle protein	15	1.95E-08	CG34461, CPR76BB, CCP84AD, CPR5C, CPR66D, CG10625, CPR49AH, CPR49AC, CPR62BC,	7.86E-06

Enrichment Score: 5.473	structural constituent of cuticle	16	2.69E-08	CPR49AA, CPR62BB, CPR56F, CPR50CB, CG1136, LCP65AD	8.65E-06
	structural constituent of chitin-based cuticle	14	4.61E-07	CG34461, CPR76BB, CCP84AD, CPR5C, CPR66D, CPR49AH, CG8541, CPR49AC, CPR62BC, CPR49AA, CPR62BB, CPR56F, CPR50CB, LCP65AD	1.48E-04
Annotation Cluster M Enrichment Score: 3.075	aromatic amino acid family metabolic process	6	7.13E-06	CG1461, CG11796, DDC, HN, PLE, HGO	0.006
	Tyrosine metabolism	5	6.94E-03	CG1461, CG11796, DDC, PLE, HGO	0.313
	Phenylalanine metabolism	4	1.20E-02	CG1461, CG11796, DDC, HN	0.479
Annotation Cluster N Enrichment Score: 2.625	extracellular region	22	1.99E-04	PEBIII, CHT5, CHT3, HML, OBST-B, OBST-A, ATTA, LSP1GAMMA, MMP2, MMP1, CG13830, NLAZ, EST-6, CG5756, CG8483, CG13643, NPC2A, CG30503, GNBP3, PGRP-SA, MTG, ANCE	0.023
	ChtBD2	8	5.04E-04	CG5756, CHT5, CHT3, CG13643, HML, OBST-B, OBST-A, MTG	0.038
	aminoglycan metabolic process	9	1.34E-03	CG5756, CHT5, CHT3, CG13643, HML, OBST-B, OBST-A, MTG, PGRP-SA	0.700
Annotation Cluster O Enrichment Score: 2.589	SEC14	6	1.34E-04	CG5958, NF1, CG10026, CG2663, CG3823, CG11550	0.010
	Cellular retinaldehyde-binding/triple function, C-terminal	6	4.39E-04	CG5958, NF1, CG10026, CG2663, CG3823, CG11550	0.162
	Cellular retinaldehyde binding/alpha-tocopherol transport	4	5.02E-03	CG5958, CG10026, CG2663, CG3823	0.868
Annotation Cluster P Enrichment Score: 2.060	nitrogen compound catabolic process	4	1.96E-03	CG42249, CG6106, GYC88E, PYD3	0.828
	nucleobase, nucleoside and	3	1.83E-02	CG42249, GYC88E, PYD3	1.000

	nucleotide catabolic process				
	nucleobase, nucleoside, nucleotide and nucleic acid catabolic process	3	1.83E-02	CG42249, GYC88E, PYD3	1.000
Annotation Cluster Q Enrichment Score: 1.857	JHBP	4	8.08E-03	CG14457, CG10407, CG13618, CG11852	0.464
	Odorant binding protein	4	1.58E-02	CG14457, CG10407, CG13618, CG11852	0.998
	Hormone binding	4	2.10E-02	CG14457, CG10407, CG13618, CG11852	1.000
Annotation Cluster R Enrichment Score: 1.678	contractile fiber	4	5.12E-03	MF, MP20, ZASP66, FLN	0.446
	sarcomere	3	2.88E-02	MF, ZASP66, FLN	0.965
	myofibril	3	3.25E-02	MF, ZASP66, FLN	0.978
Annotation Cluster S Enrichment Score: 1.570	hormone metabolic process	4	1.43E-02	EST-6, DESAT1, 7B2, CYP4AC2	1.000
	regulation of hormone levels	4	1.43E-02	EST-6, DESAT1, 7B2, CYP4AC2	1.000
	secondary metabolic process	4	9.51E-02	EST-6, DESAT1, HML, CYP4AC2	1.000
Annotation Cluster T Enrichment Score: 1.532	oxidation reduction	21	9.21E-04	DESAT1, IDH, CG17374, HN, CG17221, CYP9F2, CG31549, CG31075, CYP6A2, CG3609, PLE, CG9514, CYP4C3, CG11796, CYP6G2, CG9914, GPDH, CYP4G15, CG3523, HGO, CYP4AC2	0.563
	iron	11	1.18E-03	DESAT1, CYP6G2, HN, CYP9F2, CYP4G15, GYC88E, CYP6A2, PLE, HGO, CYP4C3, CYP4AC2	0.157
	Monooxygenase	8	1.72E-03	CYP6G2, HN, CYP9F2, CYP4G15, CYP6A2, PLE, CYP4C3, CYP4AC2	0.220
Annotation Cluster U Enrichment Score: 1.483	dendrite	4	5.90E-03	MTT, ARMI, MMP2, MMP1	0.493
	neuron projection	4	4.37E-02	MTT, ARMI, MMP2, MMP1	0.994
	cell projection	5	1.38E-01	MTT, NHA1, ARMI, MMP2, MMP1	1.000
Annotation Cluster V Enrichment Score: 1.437	Sugar transporter, conserved site	7	3.44E-04	CG6006, CG10960, CYP4G15, CG3168, CG1213, SUT4, CG6034	0.130
	General substrate transporter	4	2.70E-02	CG42269, CG10960, CG1213, SUT4	1.000
	Sugar/inositol	3	3.04E-	CG10960, CG1213, SUT4	1.000

	transporter		02		
Annotation Cluster W Enrichment Score: 1.326	aromatic amino acid family metabolic process	6	7.13E-06	CG1461, CG11796, DDC, HN, PLE, HGO	0.006
	biogenic amine biosynthetic process	3	2.96E-02	DDC, HN, PLE	1.000
	cellular amino acid derivative biosynthetic process	3	5.42E-02	DDC, HN, PLE	1.000
Annotation Cluster X Enrichment Score: 1.130	contractile fiber	4	5.12E-03	MF, MP20, ZASP66, FLN	0.446
	striated muscle cell differentiation	3	2.58E-01	MF, MP20, FLN	1.000
	muscle cell differentiation	3	3.08E-01	MF, MP20, FLN	1.000
Annotation Cluster Y Enrichment Score: 1.958	phagocytosis, engulfment	6	2.89E-03	PXD, PI3K68D, HLH106, GEL, POLO, SSE	0.718825
	phagocytosis	6	3.61E-03	PXD, PI3K68D, HLH106, GEL, POLO, SSE	0.795779
	membrane invagination	6	1.11E-02	PXD, PI3K68D, HLH106, GEL, POLO, SSE	0.992616

Annotation Clusters A-F Enrichment of DE genes between seasonal forms

Annotation Clusters G-J Enrichment of DE genes between sexes

Annotation Clusters K Enrichment of DE genes between sexes

Annotation Clusters L-X Enrcihment of DE genes between female seasonal forms

Annotation Cluster Y Enrichment of DE genes between male seasonal forms

Table S1.4. Full Data frame for Trinity contigs. Column 1 lists contig IDs assigned by Trinity. Column 2 has the length of each contig. Columns 3-14 contain the number of raw reads mapping to each contig. Columns 15-17 have *Drosophila* BLAST information such as matching Flybase symbol, E-value, and associated gene name. Columns 18-22 contain ones (1) for genes differentially expressed by season, sex, interaction, and season within females and males, zeros (0) indicate no differential expression due to the corresponding contrast.

https://academic.oup.com/mbe/article/33/1/79/2579302#supplementary-data

Table S1.5. FDR correction of two-way ANOVAs for eye development, phototransduction, and eye pigment genes.

Gene ID	Associated Gene	Sex P-value	Season P-value	Interaction P-value	Sex Q-value	Season Q-value	Interaction Q-value
comp53821_c0	*a*	0.358	0.017	0.418	0.826	0.095	0.955
comp33662_c0	*Abl*	0.578	0.785	0.832	0.950	0.677	0.991
comp53416_c0	*Afti*	0.546	0.003	0.549	0.939	0.083	0.955
comp51727_c0	*Alk*	0.335	0.754	0.679	0.826	0.671	0.987
comp40335_c0	*Amun*	0.880	0.170	0.950	0.980	0.359	0.991
comp44315_c0	*aop*	0.032	0.014	0.017	0.406	0.091	0.690
comp53898_c3	*aos*	0.070	0.550	0.710	0.508	0.596	0.987
comp45417_c0	*AP-1-2beta*	0.818	0.904	0.450	0.980	0.686	0.955
comp50375_c0	*AP-1gamma*	0.941	0.848	0.703	0.980	0.679	0.987
comp50062_c0	*AP-1mu*	0.916	0.575	0.072	0.980	0.596	0.690
comp48915_c1	*AP-1sigma*	0.001	0.035	0.527	0.132	0.159	0.955
comp47372_c0	*Arf102F*	0.603	0.633	0.695	0.953	0.626	0.987
comp32023_c0	*Arf79F*	0.350	0.530	0.490	0.826	0.596	0.955
comp46034_c0	*arr*	0.965	0.530	0.439	0.984	0.596	0.955
comp40258_c0	*Arr1*	0.790	0.439	0.343	0.980	0.544	0.955
comp48894_c0	*Arr2*	0.422	0.250	0.039	0.901	0.424	0.690
comp309785_c0	*Atx2*	0.374	0.960	0.050	0.834	0.702	0.690
comp257062_c0	*Awh*	0.764	0.150	0.650	0.980	0.346	0.987
comp22815_c0	*Axn*	0.209	0.356	0.734	0.748	0.508	0.991
comp27745_c0	*bab1*	0.513	0.869	0.530	0.939	0.679	0.955
comp53720_c1	*bab2*	0.840	0.001	0.888	0.980	0.058	0.991
comp46004_c0	*babo*	0.125	0.912	0.102	0.668	0.686	0.690
comp54026_c0	*bchs*	0.615	0.374	0.679	0.953	0.508	0.987
comp37324_c0	*B-H1*	0.187	0.787	0.524	0.730	0.677	0.955
comp39182_c0	*Blos1*	0.070	0.688	0.693	0.508	0.633	0.987
comp46845_c0	*boi*	0.655	0.036	0.306	0.974	0.159	0.955
comp48526_c0	*bun*	0.970	0.220	0.480	0.985	0.396	0.955
comp51659_c0	*Bx42*	0.289	0.013	0.195	0.782	0.091	0.881
comp50623_c0	*cac*	0.224	0.263	0.767	0.765	0.431	0.991
comp51976_c0	*Calx*	0.343	0.394	0.399	0.826	0.513	0.955
comp52742_c0	*capt*	0.477	0.074	0.940	0.925	0.235	0.991
comp52427_c0	*car*	0.233	0.122	0.650	0.765	0.330	0.987
comp53520_c0	*caz*	0.275	0.907	0.940	0.765	0.686	0.991
comp42286_c0	*CdsA*	0.219	0.260	0.359	0.765	0.431	0.955
comp47508_c0	*CG11426*	0.931	0.385	0.009	0.980	0.511	0.690
comp41553_c0	*CG7650*	0.250	0.794	0.387	0.765	0.677	0.955
comp53323_c1	*Chc*	0.535	0.962	0.298	0.939	0.702	0.955
comp48956_c0	*Chi*	0.830	0.700	1.000	0.980	0.634	1.000

comp53962_c0	*chp*	0.094	0.270	0.416	0.578	0.438	0.955
comp44157_c1	*Cib2*	0.850	0.427	0.850	0.980	0.537	0.991
comp23280_c0	*cindr*	0.870	0.553	0.444	0.980	0.596	0.955
comp31885_c0	*CkIIalpha*	0.502	0.639	0.706	0.939	0.626	0.987
comp4258_c0	*cl*	0.063	0.132	0.090	0.508	0.343	0.690
comp55875_c0	*cn*	0.267	0.493	0.444	0.765	0.594	0.955
comp48378_c0	*cno*	0.761	0.407	0.350	0.980	0.522	0.955
comp42609_c0	*crb*	0.780	0.220	0.450	0.980	0.396	0.955
comp48397_c0	*cry*	0.011	0.389	0.975	0.203	0.511	0.991
comp49784_c0	*csw*	0.483	0.006	0.800	0.925	0.091	0.991
comp50422_c0	*Dab*	0.081	0.128	0.600	0.528	0.339	0.982
comp28597_c0	*dac*	0.457	0.004	0.401	0.925	0.083	0.955
comp43170_c0	*Dad*	0.317	0.020	0.538	0.815	0.104	0.955
comp54074_c1	*dally*	0.940	0.210	0.210	0.980	0.388	0.881
comp39785_c0	*dan*	0.120	0.380	0.400	0.658	0.508	0.955
comp52105_c0	*DhpD*	0.694	0.569	0.484	0.980	0.596	0.955
comp185677_c0	*Dl*	0.243	0.632	0.870	0.765	0.626	0.991
comp50345_c0	*dlp*	0.483	0.071	0.240	0.925	0.231	0.900
comp34503_c0	*Doa*	0.540	0.020	0.890	0.939	0.104	0.991
comp50889_c0	*dor*	0.066	0.003	0.387	0.508	0.083	0.955
comp37973_c0	*dpp*	0.780	0.530	0.940	0.980	0.596	0.991
comp45812_c0	*Dronc*	0.526	0.210	0.497	0.939	0.388	0.955
comp23908_c0	*dsh*	0.113	0.165	0.137	0.643	0.359	0.732
comp44116_c0	*Dysb*	0.657	0.012	0.761	0.974	0.091	0.991
comp41953_c0	*E(spl)mbeta-HLH*	0.170	0.017	0.510	0.730	0.095	0.955
comp40007_c0	*E(spl)mgamma-HLH*	0.025	0.010	0.230	0.363	0.091	0.881
comp30822_c0	*Egfr*	0.268	0.044	0.098	0.765	0.182	0.690
comp47535_c0	*elB*	0.980	0.120	0.930	0.985	0.330	0.991
comp41270_c0	*ena*	0.105	0.066	0.987	0.627	0.224	0.992
comp47249_c0	*eya*	0.040	0.062	0.900	0.451	0.221	0.991
comp24061_c0	*eyg*	0.007	0.890	0.420	0.203	0.686	0.955
comp52449_c0	*faf*	0.324	0.895	0.603	0.822	0.686	0.982
comp48254_c0	*fl(2)d*	0.329	0.188	0.799	0.825	0.371	0.991
comp52468_c0	*fng*	0.198	0.580	0.150	0.744	0.596	0.761
comp50629_c0	*futsch*	0.640	0.076	0.970	0.973	0.236	0.991
comp45465_c0	*g*	0.529	0.340	0.969	0.939	0.497	0.991
comp40588_c0	*Galphaq*	0.468	0.543	0.318	0.925	0.596	0.955
comp44242_c0	*gl*	0.036	0.011	0.134	0.425	0.091	0.732
comp50577_c0	*Gl*	0.068	0.860	0.130	0.508	0.679	0.732
comp51417_c0	*Gp150*	0.006	0.031	0.920	0.203	0.151	0.991
comp47782_c0	*Gprk1*	0.608	0.492	0.637	0.953	0.594	0.987
comp51912_c0	*Hdc*	0.136	0.440	0.080	0.673	0.544	0.690
comp35934_c0	*hipk*	0.902	0.116	0.571	0.980	0.330	0.957

comp115040_c0	*Hn*	0.935	0.068	0.079	0.980	0.225	0.690
comp30989_c0	*holn1*	0.015	0.919	0.720	0.254	0.686	0.991
comp45330_c0	*hop*	0.822	0.873	0.888	0.980	0.679	0.991
comp48534_c0	*hth*	0.005	0.137	0.916	0.203	0.343	0.991
comp53900_c0	*hyd*	0.527	0.307	0.546	0.939	0.471	0.955
comp53699_c0	*inaD*	0.169	0.344	0.533	0.730	0.497	0.955
comp52457_c0	*inaE*	0.789	0.719	0.560	0.980	0.647	0.955
comp95479_c0	*jeb*	0.780	0.850	0.210	0.980	0.679	0.881
comp37821_c0	*jumu*	0.781	0.141	0.472	0.980	0.343	0.955
comp52768_c0	*kay*	0.001	0.015	0.110	0.132	0.091	0.711
comp39565_c0	*klar*	0.423	0.000	0.040	0.901	0.029	0.690
comp47723_c0	*Kr*	0.980	0.054	0.890	0.985	0.207	0.991
comp35250_c0	*Kr-h1*	0.895	0.320	0.780	0.980	0.486	0.991
comp54003_c0	*kto*	0.870	0.790	0.790	0.980	0.677	0.991
comp44627_c0	*L*	0.170	0.790	0.620	0.730	0.677	0.987
comp53339_c0	*l(2)gd1*	0.740	0.240	0.890	0.980	0.424	0.991
comp23468_c0	*Lim1*	0.904	0.870	0.340	0.980	0.679	0.955
comp106127_c0	*LIMK1*	0.150	0.941	0.223	0.725	0.697	0.881
comp33833_c0	*lin-52*	0.076	0.012	0.250	0.528	0.091	0.923
comp43713_c0	*lt*	0.960	0.014	0.057	0.984	0.091	0.690
comp45121_c0	*lz*	0.610	0.570	0.940	0.953	0.596	0.991
comp46432_c0	*mam*	0.690	0.248	0.749	0.980	0.424	0.991
comp46826_c0	*Mbs*	0.888	0.520	0.477	0.980	0.596	0.955
comp47516_c0	*mbt*	0.925	0.562	0.520	0.980	0.596	0.955
comp42473_c0	*Mcm10*	0.302	0.187	0.798	0.796	0.371	0.991
comp43328_c0	*mib1*	0.999	0.615	0.738	0.999	0.623	0.991
comp48375_c0	*msk*	0.264	0.054	0.205	0.765	0.207	0.881
comp53787_c0	*N*	0.888	0.152	0.449	0.980	0.346	0.955
comp53658_c0	*Nckx30C*	0.639	0.489	0.093	0.973	0.594	0.690
comp50250_c0	*nct*	0.226	0.327	0.299	0.765	0.492	0.955
comp47901_c0	*nej*	0.858	0.580	0.330	0.980	0.596	0.955
comp49707_c0	*neur*	0.433	0.602	0.501	0.906	0.614	0.955
comp52323_c0	*Nf-YB*	0.613	0.080	0.198	0.953	0.242	0.881
comp52206_c0	*ninaB*	0.741	0.146	0.098	0.980	0.343	0.690
comp53950_c0	*ninaC*	0.312	0.680	0.079	0.812	0.633	0.690
comp50825_c0	*nito*	0.165	0.690	0.330	0.730	0.633	0.955
comp34688_c0	*nmo*	0.686	0.800	0.359	0.980	0.678	0.955
comp47858_c0	*noc*	0.080	0.250	0.320	0.528	0.424	0.955
comp53566_c0	*norpA*	0.294	0.852	0.300	0.785	0.679	0.955
comp293520_c0	*Notum*	0.596	0.276	0.690	0.953	0.442	0.987
comp49074_c0	*oc*	0.064	0.950	0.167	0.508	0.700	0.827
comp46670_c0	*ogre*	0.880	0.666	0.690	0.980	0.631	0.987
comp41644_c0	*or*	0.131	0.023	0.680	0.673	0.116	0.987

comp37571_c0	*p*	0.009	0.191	0.553	0.203	0.371	0.955
comp43032_c0	*pAbp*	0.260	0.059	0.068	0.765	0.215	0.690
comp44806_c0	*Pallidin*	0.195	0.243	0.101	0.744	0.424	0.690
comp53697_c0	*PDZ-GEF*	0.570	0.015	0.380	0.950	0.091	0.955
comp53998_c0	*peb*	0.114	0.014	0.119	0.643	0.091	0.722
comp42060_c0	*pelo*	0.258	0.013	0.320	0.765	0.091	0.955
comp252943_c0	*pie*	0.176	0.360	0.090	0.730	0.508	0.690
comp50793_c0	*Pka-C1*	0.510	0.560	0.860	0.939	0.596	0.991
comp51792_c0	*Pld*	0.134	0.045	0.112	0.673	0.182	0.711
comp48267_c0	*pn*	0.414	0.575	0.359	0.901	0.596	0.955
comp53261_c0	*pnt*	0.650	0.698	0.180	0.974	0.634	0.870
comp52283_c1	*pnut*	0.255	0.850	0.680	0.765	0.679	0.987
comp47339_c0	*porin*	0.275	0.343	0.770	0.765	0.497	0.991
comp48322_c0	*Pph13*	0.094	0.064	0.226	0.578	0.222	0.881
comp53227_c0	*ptc*	0.950	0.840	0.480	0.980	0.679	0.955
comp48778_c0	*Pu*	0.642	0.652	0.036	0.973	0.629	0.690
comp209378_c0	*pwn*	0.541	0.204	0.810	0.939	0.386	0.991
comp30611_c0	*pygo*	0.570	0.560	0.930	0.950	0.596	0.991
comp54002_c0	*Rab11*	0.426	0.375	0.565	0.901	0.508	0.956
comp51205_c0	*Rabex-5*	0.370	0.180	0.080	0.834	0.371	0.690
comp51274_c0	*rap*	0.580	0.980	0.760	0.950	0.708	0.991
comp31560_c0	*Ras85D*	0.187	0.850	0.360	0.730	0.679	0.955
comp48439_c0	*RASSF8*	0.950	0.804	0.704	0.980	0.678	0.987
comp51464_c0	*rb*	0.477	0.534	0.148	0.925	0.596	0.761
comp53968_c0	*rdgA*	0.030	0.857	0.876	0.406	0.679	0.991
comp51529_c0	*rdgB*	0.066	0.156	0.966	0.508	0.350	0.991
comp53620_c0	*rdgC*	0.667	0.038	0.422	0.974	0.163	0.955
comp51219_c3	*Rh4*	0.540	0.563	0.101	0.939	0.596	0.690
comp47129_c0	*Rh5*	0.246	0.013	0.102	0.765	0.091	0.690
comp40442_c0	*Rh6*	0.664	0.922	0.084	0.974	0.686	0.690
comp44769_c0	*RhoGAP5A*	0.477	0.185	0.660	0.925	0.371	0.987
comp38351_c0	*rn*	0.930	0.296	0.774	0.980	0.459	0.991
comp238469_c0	*rno*	0.210	0.290	0.300	0.748	0.456	0.955
comp42595_c0	*RpL8*	0.730	0.680	0.840	0.980	0.633	0.991
comp52812_c0	*rst*	0.011	0.646	0.976	0.203	0.628	0.991
comp53510_c0	*ry*	0.378	0.875	0.334	0.834	0.679	0.955
comp44128_c0	*S*	0.747	0.793	0.761	0.980	0.677	0.991
comp49717_c0	*santa_maria*	0.855	0.980	0.424	0.980	0.708	0.955
comp43408_c0	*sca*	0.378	0.012	0.312	0.834	0.091	0.955
comp39552_c1	*scrib*	0.180	0.081	0.043	0.730	0.242	0.690
comp49541_c0	*se*	0.170	0.411	0.837	0.730	0.522	0.991
comp38370_c0	*sens*	0.840	0.636	0.605	0.980	0.626	0.982
comp5132_c0	*Ser*	0.178	0.195	0.678	0.730	0.374	0.987

comp52156_c0	*shakB*	0.009	0.001	0.047	0.203	0.058	0.690
comp42197_c0	*shtd*	0.205	0.411	0.892	0.748	0.522	0.991
comp29860_c0	*skd*	0.070	0.005	0.667	0.508	0.091	0.987
comp47008_c0	*Skp2*	0.570	0.170	0.480	0.950	0.359	0.955
comp53404_c0	*smo*	0.520	0.168	0.670	0.939	0.359	0.987
comp49096_c0	*so*	0.580	0.120	0.590	0.950	0.330	0.982
comp46534_c0	*Socs36E*	0.349	0.760	0.200	0.826	0.672	0.881
comp12304_c0	*spen*	0.852	0.659	0.900	0.980	0.629	0.991
comp53559_c3	*Src42A*	0.691	0.621	0.479	0.980	0.625	0.955
comp41365_c0	*ssh*	0.049	0.032	0.090	0.497	0.151	0.690
comp51633_c0	*st*	0.002	0.337	0.219	0.135	0.497	0.881
comp49385_c0	*Stat92E*	0.838	0.690	0.670	0.980	0.633	0.987
comp51699_c1	*stmA*	0.795	0.145	0.444	0.980	0.343	0.955
comp17045_c0	*sty*	0.339	0.377	0.385	0.826	0.508	0.955
comp42264_c0	*Su(var)2-10*	0.935	0.370	0.946	0.980	0.508	0.991
comp52630_c0	*svp*	0.840	0.820	0.220	0.980	0.679	0.881
comp52671_c0	*Tak1*	0.830	0.188	0.007	0.980	0.371	0.690
comp53856_c0	*Ten-m*	0.951	0.143	0.986	0.980	0.343	0.992
comp51957_c0	*tio*	0.045	0.004	0.035	0.481	0.083	0.690
comp23464_c0	*toe*	0.008	0.750	0.826	0.203	0.671	0.991
comp51658_c0	*toy*	0.709	0.011	0.828	0.980	0.091	0.991
comp53126_c0	*trp*	0.854	0.056	0.121	0.980	0.209	0.722
comp3659_c0	*TrpA1*	0.828	0.367	0.958	0.980	0.508	0.991
comp53924_c0	*trpl*	0.357	0.841	0.090	0.826	0.679	0.690
comp54033_c0	*trr*	0.817	0.660	0.903	0.980	0.629	0.991
comp41466_c0	*tsr*	0.729	0.291	0.419	0.980	0.456	0.955
comp1061_c0	*unk*	0.910	0.552	0.535	0.980	0.596	0.955
comp532780_c0	*ush*	0.478	0.859	0.537	0.925	0.679	0.955
comp52829_c0	*v*	0.282	0.262	0.799	0.774	0.431	0.991
comp43453_c0	*vito*	0.603	0.560	0.125	0.953	0.596	0.725
comp52894_c0	*Vps16A*	0.236	0.990	0.768	0.765	0.711	0.991
comp46375_c0	*Vps28*	0.769	0.530	0.840	0.980	0.596	0.991
comp53666_c0	*w*	0.022	0.010	0.560	0.344	0.091	0.955
comp373630_c0	*wg*	0.901	0.919	0.265	0.980	0.686	0.955
comp44380_c0	*wts*	0.461	0.091	0.885	0.925	0.265	0.991
comp48795_c0	*X11L*	0.850	0.140	0.084	0.980	0.343	0.690

Table S1.6. Primer sequences for RT-PCR.

Target gene	Primer	Sequence
CG2663	Forward	CGGTTAACTGTCTAAAGGAGCC
CG2663	Reverse	AGGAAGCACAGAAACTAATCGG
henna	Forward	TTGTATCTCACGCCGAACTC
henna	Reverse	GAACAGAAGTACCCCATCACG
warts	Forward	TCGCTTTCAACCCACTTACC
warts	Reverse	GGTGTTGATCTGGATACTAGCG
klar	Forward	GGCGTTTTGTTCGACTTGAG
klar	Reverse	CGTGAGCACCAATACTGATCC
domeless	Forward	CGCCAACCTGACTACCATAAA
domeless	Reverse	GGATCGAGTTTGCAGTTTTGG
blue	Forward	CGCGAGTGCAAGCATCTC
blue	Reverse	CACGAATTTTCCCCAGATCCTGAA
Ef1alpha	Forward	CAACCTAGTACCATCCAAACCC
Ef1alpha	Reverse	TCCTTGAAGTTGACAGCCTTG
white	Forward	GTGACTTTGTGGCTGTTATTGG
white	Reverse	GATCTCTATCACACCTCGCTG

CHAPTER 2

Copy number variation and expression analysis reveals a non-orthologous *pinta* gene family member involved in butterfly vision

ABSTRACT

Vertebrate (CRALBP) and *Drosophila* (PINTA) proteins with a CRAL-TRIO domain transport retinal-based chromophores that bind to opsin proteins and are necessary for phototransduction. The CRAL-TRIO domain gene family is composed of genes that encode proteins with a common N-terminal structural domain. While there is an expansion of this gene family in Lepidoptera, there is no lepidopteran ortholog of *pinta*. Further, the function of these genes in lepidopterans has not yet been established. Here we explored the molecular evolution and expression of CRAL-TRIO domain genes in the butterfly *Heliconius melpomene* in order to identify a member of this gene family as a candidate chromophore transporter. We generated and searched a four tissue transcriptome and searched a reference genome for CRAL-TRIO domain genes. We expanded an insect CRAL-TRIO domain gene phylogeny to include *H. melpomene* and used 18 genomes from 4 subspecies to assess copy number variation. A transcriptome-wide differential expression analysis comparing four tissue types identified a CRAL-TRIO domain gene, *Hme CTD31*, upregulated in heads suggesting a potential role in vision for this CRAL-TRIO domain gene. RT-PCR and immunohistochemistry confirmed that *Hme CTD31* and its protein product are expressed in the retina, specifically in primary and secondary pigment cells and in tracheal cells. Sequencing of eye protein extracts that fluoresce in the ultraviolet identified Hme CTD31 as a possible chromophore binding protein. Although we found several recent duplications

and numerous copy number variants in CRAL-TRIO domain genes, we identified a single copy *pinta* paralog that likely binds the chromophore in butterflies.

INTRODUCTION

Phenotypic differences between organisms may be driven by small nucleotide changes in protein coding or regulatory regions, or by whole gene or genome duplications (Stern 2000; Hersh & Carroll 2005; Demuth et al. 2006). Gene duplications in particular are hypothesized to be an important mechanism for evolutionary change because these events give rise to new material for novelties and may facilitate the emergence of new genes (Ohno 1970; Long et al. 2003). Often, gene duplications result in pseudogenization. However, there are at least two mechanisms by which duplicated genes can remain functional regardless of redundancy: 1) in neofunctionalization, a duplicated gene develops a new function different from the ancestral gene and 2) in subfunctionalization the two paralogs each have part of the function of an ancestral gene (Lynch & Conery 2000; Long et al. 2003; Zhang 2003). Gene duplications and rearrangements have resulted in large gene families. Genes are classified as part of a gene family when they share common sequence motifs and sometimes may have related general functions (Henikoff 1997).

Lineage-specific gene family expansions are hypothesized to be a mechanism by which eukaryotic species can adapt and diversify (Lespinet et al. 2002). In support of this, studies in mammals suggest that changes to the size of large gene families are likely arising through lineage specific gene loss or gain rather than by changes in gene number at branch sites (Demuth et al. 2006). Gene families that are subject to expansions or reductions have a wide variety of functions, including immunity and sensory perception (Cooper et al. 2007; Dopman & Hartl 2007; Conrad et al. 2010). Chemosensory genes in particular have been

widely studied in a number of species and have been found to vary in copy number between and within species (Nei et al. 2008; Nozawa & Nei 2008). Copy number variation (CNV) is a DNA segment 1 kb or longer whose copy number varies between individuals, as a result of recent gene duplications or deletions (Stranger et al. 2007). Insects have been studied for CNV by focusing on gene families with lineage-specific duplications; these genes are candidates for CNVs (Zhang 2003). As an example, the butterfly *Heliconius melpomene* and the pea aphid *Acyrthosiphon pisum* both have lineage-specific gene expansions and CNV of olfactory and gustatory receptors correlated with host plant specialization (Briscoe et al. 2013; Duvaux et al. 2014).

The CRAL-TRIO domain gene family is another family that is evolving by lineage-specific duplication in insects and has undergone an expansion in Lepidoptera (Smith & Briscoe 2015) (moths and butterflies; Smith and Briscoe 2015). Lepidoptera thus have almost twice as many CRAL-TRIO domain genes relative to other insects (Smith & Briscoe 2015). The lineage-specific duplications of this gene family make it a candidate to study for CNV (Zhang 2003). Furthermore, the specific functions of the members of this family remain unknown, with one or two exceptions. The CRAL-TRIO domain is an N-terminal structural region, approximately 170 amino acids long, common to several proteins that bind and transport tocopherols (Panagabko et al. 2003; Sigrist et al. 2013). The CRAL-TRIO domain gene family includes a cellular retinaldehyde-binding protein (CRALBP) which is essential to vertebrate vision due to its function in chromophore transport (Wu et al. 2006). The visual pigment chromophore is derived from vitamin-A. In photoreceptor cells, opsin proteins bind a chromophore molecule (in humans 11-*cis*-retinal and in butterflies 11-*cis*-3-hydroxyretinal) to form rhodopsin. Rhodopsin initiates the phototransduction

cascade when photon absorption changes the chromophore configuration from 11-*cis* to all-*trans* (von Lintig et al. 2010). In humans, mutations to CRAL-TRIO domain genes result in a variety of retinal and neurological diseases (Maw et al. 1997; Bomar et al. 2003; Min et al. 2003). Mutations in RLBP1, the gene encoding CRALBP in humans, results in retinitis pigmentosa (Maw et al. 1997) and mutations in a gene encoding αTTP results in ataxia with vitamin E deficiency (AVED) (Min et al. 2003). Moreover, mutations in human *Atcay*, a CRAL-TRIO domain containing gene, are associated with Cayman ataxia, and a mouse homolog of *atcay* causes ataxia and dystonia in jittery mice (Bomar et al. 2003).

In insect genomes, CRAL-TRIO domain genes are numerous, however, their function remains largely unexplored except for *prolonged depolarization afterpotential is not apparent* (*pinta*). PINTA in *Drosophila* is a CRAL-TRIO domain protein belonging to the SEC14 superfamily that, similar to CRALBP, shuttles the chromophore from retinal pigment cells to photoreceptor cells (Wang & Montell 2005). PINTA protein is required for the biosynthesis of rhodopsin. *Drosophila* with mutated *pinta* genes have low expression of Rh1, the protein component of the light-sensitive rhodopsin found in R1-6 photoreceptor cells (Wang & Montell 2005). Similarly, another member of the SEC14 superfamily squid RALBP functions in retinal binding in cephalopods (Ozaki et al. 1994; Speiser et al. 2014). Although there is an expansion of CRAL-TRIO domain genes in Lepidoptera, no *pinta* ortholog has been found in this group. The functions of CRALBP and PINTA suggest that a distinct CRAL-TRIO domain protein might be serving an essential role in lepidopteran visual systems by transporting the chromophore.

Presently most of our knowledge about photoreceptor determination, phototransduction, and chromophore transport comes from studies in *Drosophila*.

However, a recent analysis of 80 vision genes in the *Manduca sexta* genome (Kanost et al. 2016) found that at least four gene families involved in photoreceptor differentiation pathways have undergone lepidopteran-specific gene duplications including *corkscrew*, *embryonic lethal/abnormal vision*, rhabdomeric opsins, and genes encoding CRAL-TRIO domain containing proteins. Since CRAL-TRIO domain genes have undergone an expansion in Lepidoptera and their functions in other organisms suggest a role in vision, it is worth exploring the function of these genes in a butterfly species. *Heliconius melpomene* provides a good system in which to investigate the evolution of CRAL-TRIO domain genes due to the availability of a reference genome and a growing collection of resequenced genomes (Martin et al. 2013; Davey et al. 2016). In addition, we have generated RNA-Seq data from *Heliconius melpomene* tissues with which to investigate the expression of the CRAL-TRIO domain genes.

Here, we aim to 1) characterize the molecular evolution of the CRAL-TRIO domain gene family and to 2) identify a candidate gene for chromophore transport in butterflies. We used RNA-Seq data from *H. melpomene* head, antennae, legs and mouth parts to make a *de novo* transcriptome assembly from which to identify CRAL-TRIO domain gene transcripts. We also investigated the reference genome to search for any CRAL-TRIO domain genes that may be found in the genome but not expressed in the tissues we sampled. We found support for the expansion of the CRAL-TRIO domain gene family in butterflies by identifying 43 CRAL-TRIO domain genes in the *H. melpomene* genome comparable to the 42 found in *Manduca sexta* (Smith & Briscoe 2015). We also investigated 18 resequenced *H. melpomene* genomes (Martin et al. 2013) for structural variation (specifically CNV) and found that 32 of the 43 genes in the reference genome had either a

large duplication or deletion in at least one of the resequenced genomes. Further, to identify a CRAL-TRIO domain gene functioning in vision, we did a differential expression analysis between tissue types and found one CRAL-TRIO domain gene (*Hme CTD31*) that is upregulated in head tissue. RT-PCR and immunohistochemistry shows that *Hme CTD31* is expressed in the compound eye and not the brain, and Hme CTD31 is localized to the retinal pigment and trachea cells making it a candidate chromophore binding protein. We used mass spectrometry to sequence eye proteins associated with an ultraviolet fluorescing pigment and found a match for our CRAL-TRIO domain protein Hme CTD31. These various lines of evidence suggest that we have found a CRAL-TRIO domain gene that binds the chromophore in butterflies.

MATERIALS AND METHODS

CRAL-TRIO domain gene phylogeny

A phylogeny from Smith and Briscoe (2015) (Smith & Briscoe 2015) was expanded by adding homologs of the CRAL-TRIO domain gene family found in *H. melpomene*. Smith and Briscoe identified CRAL-TRIO domain genes from the genomes of *Manduca sexta*, *Danaus plexippus*, *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Tribolium castaneum* and *Bombyx mori* (Smith & Briscoe 2015). To expand this repertoire, we used BLAST+ (Basic Local Alignment Search Tool) (Camacho et al. 2009) to identify CRAL-TRIO domain gene homologs in a *de novo* transcriptome of *H. melpomene rosina*. These contig sequences were extracted and added to the alignment. Contig nucleotide sequences were translated and curated in MEGA by finding the correct reading frame from start to stop codon. Sequences with missing homologs were blasted against the *H. melpomene melpomene* reference genome v. 2 (Davey et al. 2016), from which additional sequences

were recovered. Manual annotations of the genes not included in the transcriptome and not annotated in the reference genome were done by extracting the nucleotide sequence around the area where there was a BLAST hit to a CRAL-TRIO domain gene. The extracted nucleotides were annotated and translated in AUGUSTUS (Stanke & Morgenstern 2005) and aligned to a BLAST output of the genome to correct the sequence. 215 amino acid sequences were aligned using MUSCLE (Edgar 2004) with default settings, and this alignment was then modified manually. A Bayesian phylogenetic tree was made using MrBayes (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003; Ronquist et al. 2011) with a BLOSUM62 (Henikoff & Henikoff 1992) model for 1,000,000 generations. The phylogeny was color coded using iTOL (Letunic & Bork 2016).

Structural Variation

To detect copy number variation (CNV) in these genes, we aligned reads for 18 resequenced *H. melpomene* genomes generated by Martin et al. (Martin et al. 2013), European Nucleotide Archive: ERP002440. Read mapping to the reference genome for 4 subspecies (6 *H. melpomene melpomene*, 4 *H. melpomene rosina*, 4 *H. melpomene amaryllis*, 4 *H. melpomene aglaope*) were performed using bwa (Li & Durbin 2009), and samtools was used to index and sort the files (Li et al. 2009). Pindel was used to examine mapping results to detect structural variation (Ye et al. 2009). Pindel looks for read pairs for which one read maps uniquely to the genome while the other read is unmapped to determine the structural breakpoint and direction of unmapped reads (Ye et al. 2009).

RNA library preparation

We extracted RNA from whole heads (excluding antennae and mouth parts) of three male and three female *H. melpomene* butterflies. We also extracted RNA from the head,

antennae, legs, and mouth parts (lapial palps + proboscis) of one male and one female *H. melpomene* specimen to increase our biological replicates to n=4. Butterflies were placed in -80°C and stored until RNA extraction. RNA was extracted using TRIzol (Life Technologies, Grand Island, NY) and purified using a NucleoSpin RNA II kit (Macherey-Nagel, Bethlehem, PA). Purified RNA was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY) and quality checked using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). A TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, CA) was used to prepare sequencing libraries. Libraries with distinct adapter sequences were quantified, quality checked, normalized and pooled according to their concentrations. Pooled libraries were run on a 2% agarose gel. A Geneclean III kit (MP Biomedical, Santa Ana, CA) was used to recover DNA from the gel (~240-600 bp), and Agencourt AMPure XP (Beckman Coulter, Brea, CA) beads were used for a second purification. Sequencing was conducted at the UCI Genomics High-Throughput Facility using a HiSeq 2500 (Illumina, San Diego, CA), paired end 100-cycle sequence run.

Assembly and read-mapping

RNA-Sequencing data were obtained for three *H. melpomene* males and three female antennae, legs, and mouth parts from a previous RNA-Seq study (ArrayExpress: E-MTAB-1500) (Briscoe et al. 2013) . We created eight new head libraries from four males and four females. In addition, we made a new antennae, legs, and mouth parts library for one *H. melpomene* male and one female. The raw sequencing data for the 14 new libraries were deposited in ArrayExpress archive under E-MTAB-6249 and E-MTAB-6342. All libraries were parsed using custom perl and python scripts. A *de novo* transcriptome assembly was constructed using Trinity (Grabherr et al. 2011; Haas et al. 2013) by including one library

per tissue type (head, legs, antennae, mouthparts) for one male and one female, 8 libraries total. We made a *de novo* assembly because the CRAL-TRIO domain genes were not all annotated in the genome and a transcriptome recovered more sequences that were complete. The reference transcriptome was deposited in Dryad under doi:10.5061/dryad.857n9. Each sequenced library was then mapped back to the reference assembly using RSEM (Li & Dewey 2011) from which we extracted raw read count data, FPKM (Fragments Per Kilobase of exon per Million fragments mapped), and TPM (Transcripts Per Kilobase Million). FPKM was further normalized using NOISeq (Tarazona et al. 2011).

Since some of the CRAL-TRIO domain genes were not recovered in the transcriptome, we manually annotated these genes and read mapped each library as described above to the nucleotide sequences of the 43 genes. TPM expression was scaled to the values of whole-transcriptome analysis. We then used two-way ANOVAs to test if these genes varied by sex, tissue type, or sex and tissue type interaction

edgeR

We performed differential gene expression analysis for all Trinity assembled contigs using edgeR (Robinson & Smyth 2007, 2008; Mccarthy et al. 2012; Robinson et al. 2010). To analyze genes differentially expressed by tissue type, we did pairwise comparisons of head vs. antennae, head vs. legs, and head vs. mouth parts using a generalized linear model with terms for tissue, sex, the interaction of sex and tissue and included a term to correct for batch effects of sequencing on different lanes (~batch + tissue + sex + sex*tissue). Each analysis included filtering to remove contigs expressed at less than 1 count per million (CPM) for at least 4 groups, and between sample normalization using a trimmed mean of

the log expression ratios (TMM) (Robinson & Oshlack 2010). Contigs were considered significantly differentially expressed when the false discovery rate (FDR) was less than 0.05 and the log fold change (logFC) was greater than 1. We did FDR corrections using the qvalue package and using a Bonferroni correction (Storey & Tibshirani 2003; Dabney & Storey 2013).

RT-PCR

To localize where in the head the candidate gene was expressed, we performed reverse transcription polymerase chain reaction (RT-PCR) using RNA from a single individual male and female *H. melpomene* antennae, retina, and brain tissue. Animals were sacrificed a day after eclosion by squeezing the thorax. Heads were dissected in petri dishes in Ringer's solution, the retina and brain tissue were transferred to 1.7 ml microtubules on ice. Total RNA was extracted from these tissues using TRIzol (Life Technologies, Grand Island, NY) and quantified using a Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY). RNA was treated with DNAse I (Fisher Scientific, Pittsburgh, PA). Primers were designed using Primer3 (Table S2.1) (Koressaar & Remm 2007; Untergasser et al. 2012). Each 25 µl reaction had 2.5 µl Choice PCR buffer (Denville Scientific, South Plainfield, NJ), 2.5 µl dNTPs (2 mM), 0.5 µl Choice-Taq Blue (Denville Scientific, South Plainfield, NJ), 0.5 µl (1:20 diluted) SuperScript® II Reverse Transcriptase (Life Technologies, Grand Island, NY), 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 18 µl H2O and 1 µl RNA (12 µg/mL). The PCR reaction consisted of 42°C for 30 min, 20 cycles of (95°C for 30 sec, 55°C for 30 sec, and 68°C for 55 sec), 68°C for 7 min, 4°C hold. We visualized amplification by running the PCR products on a 2% agarose gel.

Immunohistochemistry

An antibody against the peptide N-CLRPGKPTNYDELFGID-C of the *Heliconius melpomene* CTD31 was generated in chicken and immunoaffinity purified (New England Peptide, Gardner, MA). We also used a rabbit antibody against the nymphalid *Limenitis astyanax* LWRh opsin sequence (Frentiu et al. 2007) to label LWRh expressing cells in *H. melpomene* (McCulloch et al. 2016). Eyes were fixed, sucrose protected, cryosectioned and immunolabeling was performed as described in McCulloch et al. (2016) (McCulloch et al. 2016). Slides were placed in 100% ice-cold acetone for 5 minutes, then washed 3 x 10 minutes in 0.1 M Phosphate-buffered saline (PBS). Slides were then placed in 0.5% sodium dodecyl sulfate in 0.1 M PBS for 5 minutes. Each slide was blocked for one hour at room temperature using 8% (v/v) normal goat serum, and 0.3% Triton X-100 in 0.1 M PBS. Slides were incubated with 1:75 chicken anti-CTD31 and 1:15 rabbit anti-LWRh antibodies in blocking solution overnight at 4°C. Slides were washed 3 x 10 minutes in 0.1 M PBS and then incubated with 1:1000 goat anti-chicken Alexafluor 488 and 1:500 goat anti-rabbit Alexafluor 555 secondary antibodies in blocking solution for two hours at room temperature in the dark. Slides were washed once more 3 x 10 minutes in 0.1 M PBS in the dark. Slides were stored for imaging by coverslipping with Aqua Poly/Mount (Polysciences, Inc. Cat. # 18606). Image stacks were taken using a Zeiss LSM700 Confocal Microscope under 20x objective at the UC Irvine Optical Biology Core Facility. Maximum Intensity projections and two-channel composites were generated using Fiji. Brightfield images were taken using untreated sections and were viewed with epifluorescence microscopy using a Zeiss Axioskop 2 under a 20x lens. Images were taken using a Canon PowerShot S5 and

associated Canon software. Contrast and brightness were adjusted for clarity using Adobe Photoshop and Fiji.

Western Blot and Mass Spectrometry

Butterfly heads were removed and immediately placed at -80°C until they were shipped together with an aliquot of anti-CTD31 antibody to Zyagen (San Diego, CA) overnight on dry ice. Immunoblotting was performed by Zyagen. Proteins were extracted by mechanical homogenization in protein lysis buffer and estimated protein concentration using a BCA kit. Total protein was fractionated through 2 large gels (SDS-PAGE) at different concentrations (20, 40, 60, and 80 μg each gel). Protein from the two gels was then transferred to Polyvinylidene fluoride (PVDF) membranes. The membrane of the first gel was blocked with 5% milk in TBST for 2 hours, then incubated with primary antibody anti-CTD31 at a concentration of 1:100 at 4°C overnight. Membrane was washed 3 times in TBST then incubated with secondary antibody (anti-chicken-peroxidase antibody) from Jackson ImmunoResearch (West Grove, PA) at a concentration of 1:5000 for 1 hour. After several washes, membrane was incubated for 5 min with chemiluminescence substrate. Two major protein bands were observed around 35 KD.

To visualize which protein may be interacting with the chromophore, 8 aliquots (50 ug each) of butterfly head proteins were fractionated through large native gel by Zyagen (San Diego, CA). One lane was cut and visualized on UV light to locate bands that fluoresce. Gel pieces containing three protein bands were collected in 15-ml tubes and were shipped to UC Irvine. The samples were immediately transferred to a Proteomics & Mass Spectrometry Facility in the school of Biological Sciences (Irvine, CA) for nano LC-MS/MS mass spectrometry using an LTQ Velos Pro mass analyzer (Thermo-Fisher). The resulting

peaklists were compared against our translated transcriptome along with a database of common contaminants using Mascot 2.5 to score (Matrix Science, Boston, Massachusetts). Mascot scores are the probability that the ion score of the experimental data match the ion scores of the database sequence; protein scores greater than 67 are significant ($p<0.05$).

RESULTS AND DISCUSSION

Phylogeny and chromosomal location

We identified a total of 43 CRAL-TRIO domain genes (*Hme CTD*) in the *H. melpomene* reference genome and 28 of them were recovered in a *de novo* assembly (Table S2.2). We found *H. melpomene* orthologs of most previously identified insect CRAL-TRIO domain genes (Smith & Briscoe 2015). We also discovered a recent duplication (*Hme CTD38* and *CTD39*) since *Heliconius* shared a common ancestor with *Danaus plexippus,* and an expansion of CRAL-TRIO domain genes (*Hme CTD16-20* and *Hme CTD24-25*; Figure 2.1). We refer to recent paralogs found in the reference genome as recent duplications; we refer to multiple duplications as an expansion, and genes with CNV are those that are duplicated or deleted in resequenced genomes compared to the reference. We named the *H. melpomene* CRAL-TRIO domain genes according to their location on scaffolds and since many genes are on similar scaffolds, we decided to map these genes on to chromosomes (Figure 2.2A). We found that all 43 genes were located on a total of 5 chromosomes and 23 of the genes were on a single chromosome, chromosome 2 (Figure 2.2A). Only one gene in this family (*Hme CTD44*) is intronless and likely arose through retrotransposition (Zhang 2003).

New genes also arise by tandem duplication which themselves arise by unequal crossing over resulting in new gene copies adjacent to each other or by segmental duplications which can be dispersed throughout the genome and experience few

recombination events (Jelesko et al. 1999; Baumgarten et al. 2003; Zhang 2003; Cannon et al. 2004). Most of the CRAL-TRIO domain genes were located in tandem suggesting that this gene family is the result of early segmental duplications and recent tandem duplications or early and recent tandem duplications with rearrangements in *H. melpomene* (Figure 2.2A). Moreover, areas of gene duplication can be hotspots for chromosomal rearrangement and might be enriched for copy number variation (CNV) (Sharp et al. 2005). In *D. melanogaster*, tandem duplications are significantly enriched near areas with CNVs (Dopman & Hartl 2007). The physical locations of CRAL-TRIO domain genes display arrays of tandem duplications making this gene family a good candidate for studying CNV (Redon et al. 2006).

Copy Number Variation

We used Pindel (Ye et al. 2009) to look for duplications and deletions 1 kb or larger (Stranger et al. 2007) of these CRAL-TRIO domain genes in resequenced genomes of four *H. melpomene* subspecies, *H. melpomene melpomene*, *H. melpomene rosina*, *H. melpomene amaryllis*, and *H. melpomene aglaope* (Martin et al. 2013). The average size of these genes including introns is 3,648 bp, coding sequences being approximately 304 amino acids long. Nine genes (*Hme CTD1-9*) were located on chromosome 1; *Hme CTD2*, *3* and *5-8*, had potential CNV in at least one of the 18 sampled genomes (Figure 2.2A). *Hme CTD2-3* were duplicated in one *H. melpomene aglaope* individual and were deleted in two genomes (*H. m. melpomene* and *H. m. aglaope*; Figure 2.2B; Table S2.3; Table S2.4). *Hme CTD5-8* were duplicated in two genomes (*H. m. melpomene* and *H. m. aglaope*; Figure 2.2B; Table S2.3), but *Hme CTD7-8* were deleted in one *H. m. amaryllis* genome (Table S2.4).

Twenty-three genes (*Hme CTD10-33*) were located on chromosome 2 with more complex patterns of CNV. *Hme CTD10-13*, *15-20*, *22*, *24-25*, and *31* were potentially duplicated in one or more resequenced genome (Figure 2.2B). Of these duplicates, *Hme CTD11-13* and *33* were duplicated in one resequenced genomes (Table S2.3). *Hme CTD10* and *22* were duplicated in 2 genomes (Table S2.3). *Hme CTD15, 20, 24* and *25* were duplicated in 3 genomes (Table S2.3). *Hme CTD19* was duplicated in 4 genomes (Table S2.3). *Hme CTD16* was duplicated in 5 genomes and *17-18* were duplicated in 7 of the 18 resequenced genomes (Table S2.3). Multiple CRAL-TRIO domain genes were also deleted in at least one resequenced genome: *Hme CTD11*, *15-23*, *25*, *27-28*, *30* and *32* (Figure 2.2A). Of these, *Hme CTD11*, *15*, *21* and *30* were deleted in one resequenced genome (Figure 2.2B; Table S2.4). *Hme CTD23*, *28*, and *32* were deleted in 2 genomes (Figure 2.2B; Table S2.4). *Hme CTD22* and *27* were deleted in 3 genomes (Figure 2.2B; Table S2.4). *Hme CTD18-20* were deleted in 4 genomes (Figure 2.2B; Table S2.4). *Hme CTD16-17* and *25* were deleted in 5 genomes (Figure 2.2B; Table S2.4). One sequence identified in the *de novo* transcriptome, *Hme CTD26*, was excluded from analysis because the translation of the mRNA contig included stop codons and BLAST results suggested it was a chimeric sequence of *Hme CTD24* and *25,* most likely due to a Trinity misassembly. In some instances, duplications and deletions are large enough to change the presence or absence of a few genes in close proximity. Genes with the most duplications/deletions were duplicated/deleted in different subspecies; this shows that there is CNV between and within subspecies.

Seven genes (*Hme CTD34-40*) were found on chromosome 4; none were duplicated but *CTD36* was deleted in one *H. m. amaryllis* and one *H. m. aglaope*, and *CTD38-39* were both deleted in one *H. m. melpomene* genome (Figure 2.2B; Table S2.4). One CRAL-TRIO

domain gene (*Hme CTD41*) was located on chromosome 12, this gene was duplicated in one *H. m. aglaope* genome (Figure 2.2B; Table S2.3). Lastly, *Hme CTD42-44* were on chromosome 21; *CTD41* was duplicated in one genome, *CTD42* was duplicated in one genome and deleted in 4 genomes, *CTD43* was duplicated in 3 and deleted in 2 genomes, and all resequenced genomes had one copy of *CTD44* (Figure 2.2B; Table S2.3; Table S2.4). To summarize, we found potential CNV in 32 of the 43 CRAL-TRIO domain genes. Intriguingly, we found no CNV in *Hme CTD31*, our candidate chromophore-binding protein (see below).

We refer to our findings of structural variation as "potential" duplications or deletions because the results were derived through bioinformatic inference which is subject to error (Emerson et al. 2008; Alkan et al. 2011). Pindel uses read mapping information in order to find paired reads in which one read maps to the reference and the other mate does not to identify break points and direction of unmapped reads (Ye et al. 2009). For a few large areas with a lot of potential structural variation, Pindel could not differentiate whether the break was due to a duplication or deletion. Although current CNV analyses are subject to error, finding replication of duplications or deletions in more than one resequenced genome as we found in some instances is evidence that these results are meaningful. We investigated the breakpoints for genes that were duplicated/deleted in multiple resequenced genomes and found that a majority of genes had similar breakpoints in at least 2 individuals (Table S2.3; Table S2.4). In addition, a different study investigated CNVs in *H. melpomene rosina* using three discovery methods and found support for duplications in the genome location of *Hme CTD5-9* and *CTD16-18* (Pinharanda et al. 2016, 2017). That study also used Pacific Biosciences (PacBio) long molecule sequencing of *H.*

melpomene and *H. cydno* to validate the findings of CNVs on chromosome 2. They found support for CNV in *Hme CTD10-12* using one discovery method and found many instances of CNVs in scaffold Hmel202006 using the three discovery methods (Pinharanda et al. 2016, 2017).

Twenty of our CRAL-TRIO domain genes were located on scaffold Hmel202006 including the genes within the *H. melpomene* expansion (*Hme CTD16-20* and *24-25*). We find the most CRAL-TRIO domain genes in tandem at a scaffold where our study and another found a large amount of CNV (Pinharanda et al. 2016, 2017). An interesting observation of CNV in this gene family was that all of the genes within the *H. melpomene* CRAL-TRIO expansion have potential CNV between individuals. In particular *Hme CTD16-20* have potential CNV in the highest number of resequenced genomes (n=9, 9, 8, 6, 5) relative to other CRAL-TRIO domain genes. These results suggest that this area in the genome could be a hotspot for structural variation potentially due to unequal crossing over because similar duplicates are located in tandem.

The adaptive significance of CNV is still under investigation. As mentioned previously, the number of chemosensory receptor genes present between and within animal species is variable (Nozawa & Nei 2008) and their distribution suggests CNV is the result of genomic drift that can lead to adaptive evolution (Nozawa & Nei 2008). In *Drosophila melanogaster*, duplications with functional sequences were found to be possibly beneficial (Dopman & Hartl 2007). CNV affects phenotypes through its direct influence on gene expression. In humans, CNV can lead to Mendelian and complex diseases by affecting gene dosage (Redon et al. 2006). The HapMap project found a substantial amount of CNV between humans, and an association analysis determined that most significant CNV-

associations had a positive correlation between gene copy number and gene expression levels (Stranger et al. 2007). Several positively selected duplication and deletion events in *D. melanogaster* have also been linked to gene expression variation (Emerson et al. 2008; Schmidt et al. 2010; Catalán et al. 2016).

Studies in *Drosophila* suggest CNV persists due to positive selection on paralogs that have tissue-specific expression (Dopman & Hartl 2007). To determine expression patterns for CRAL-TRIO domain genes we looked at gene presence and absence in the head, antennae, legs and mouth parts of male and female *H. melpomene* (n=4/sex). Here we refer to complete presence as having > 1 Transcripts Per Kilobase Million (TPM) for all replicates, partial > 1 TPM for at least one replicate but not all four, and absence as mRNA expression <1 TPM for all replicates (Figure 2.2C). Some genes varied in presence patterns between tissue types such as *Hme CTD22*, *28*, and *38* (Figure 2.2C). *Hme CTD4-9*, *12-15*, *20-21*, *23-25*, *30-35*, *39*, *41*, and *42* had different presence patterns between sexes for one or more of the tissues examined. Although patterns of gene presence or absence (Figure 2.2C) provide an idea of which genes are expressed and where, absolute and differential expression needs to be analyzed to detect potential gene functions (see below).

CNV may be one contributor to the speciation of *Heliconius*, which has undergone a radiation in Central and South America (Kozak et al. 2015; Pinharanda et al. 2017). A recent study sought to identify adaptive CNV between two sympatric hybridizing species with distinct wing patterns, *H. melpomene* and *H. cydno* (Pinharanda et al. 2017). That study found four duplications with strong signals of divergent selection: these included an odorant binding protein, a serine protease, a regulator of the cell cycle and nitrogen compound metabolic processes, and one near the gene *cortex* which regulates wing color

patterns (Nadeau et al. 2016; Pinharanda et al. 2017). The identification of an odorant binding protein supports the finding of *Heliconius* species having CNV of olfactory and gustatory receptor genes for putative host plant recognition in oviposition behavior (Briscoe et al. 2013). Divergent selection of a serine protease could be associated with *Heliconius* pollen feeding behavior (Smith et al. 2016). This raises the question as to what is the function of the CRAL-TRIO domain genes which have potential CNV between and within species.

Differential expression analysis

Members of the CRAL-TRIO domain protein family are believed to be involved in transporting hydrophobic molecules. In particular, a member of this gene family (*pinta*) transports the chromophore necessary for phototransduction in *Drosophila*, however we did not find an ortholog in Lepidoptera (Figure 2.1). To detect whether any of the CRAL-TRIO domain genes in *H. melpomene* might have this function, we did a differential expression analysis to identify CRAL-TRIO domain genes upregulated in head tissues (relative to antennae, legs, and mouth parts), potentially involved in vision. We built a reference transcriptome assembly consisting of 68,388 transcripts and 31,193 contigs with an N50 of 2,627. On average 10 million reads mapped to the transcriptome and each library averaged 79% read mapping (Table S2.5). The transcriptome was deposited in Dryad under data identifier doi: 10.5061/dryad.857n9 and the raw RNA-Seq reads were deposited in ArrayExpress archive under accession MTAB-6249 and E-MTAB-6342.

Differential gene expression analysis comparing heads vs. antennae yielded 4,868 Differentially Expressed (DE) contigs using qvalue and 1,173 using Bonferroni for false discovery rate correction (Table S2.6), 561 of these 1,173 contigs were upregulated in

heads (Table 2.1). Analysis of head vs. legs mRNA gave 6,108 DE contigs using qvalue and 1,472 using Bonferroni (Table S2.7), of these contigs 928 were upregulated in heads. Heads vs. mouth parts comparison yielded 6,176 DE contigs using qvalue and 1,486 using Bonferroni (Table S2.8), 914 of these were upregulated in heads (Table 2.1).

CRAL-TRIO domain genes expression

To find if any CRAL-TRIO domain genes were potentially upregulated in *H. melpomene* heads, we inspected our significantly DE gene list for CRAL-TRIO domain genes. By using the Bonferroni method to correct for multiple tests, only one CRAL-TRIO domain contig was upregulated in the head vs. antennae comparisons, *Hme CTD31* (Table 2.2). This gene was also upregulated across comparisons when qvalue was used to correct for multiple tests. *Hme CTD22* was upregulated in head vs. antennae and head vs. legs when using qvalue, but *Hme CTD31* was the only contig upregulated across all comparisons. In addition, when we plotted the TPM (Transcripts Per Million) for all genes across tissues, it became apparent that *Hme CTD31* is very highly expressed in male and female heads (Figure 2.3; Figure S2.1-4).

To investigate patterns of gene expression in the rest of the CRAL-TRIO domain genes, we used two-way ANOVAs to test if these genes varied by sex, tissue type, or sex and tissue type interaction (Table S2.9). We found that most genes varied by tissue type, including *Hme CTD1*, *2*, *4*, *5*, *7-14*, *21-24*, *27, 29*, *31*, *33-40*, *43*, and *44* (Table S2.9, supplementary fig. S5-7). Only two genes varied by sex *Hme CTD7* and *8* (Table S2.9; supplementary fig. S5-7).

The ANOVA analysis and the genome-wide DE analysis showed that *Hme CTD31* is a candidate pigment binding protein due to high expression in *H. melpomene* heads. The top

NCBI blastp (protein to protein alignment) results for this gene are CRAL-TRIO domain containing protein and alpha-tocopherol transport protein. We found CRAL-TRIO domain genes that were upregulated in other tissues such as *Hme CTD11* and *CTD12* in the antennae. We do not know the specific function of these genes, but it is possible that they play a role in mediating the activation of other sensory receptors. Studies identifying chemosensory proteins have found some potential sensory receptors that are similar in sequence to opsins (Troemel et al. 1995). Opsins and some chemosensory receptors, such as olfactory, gustatory and ionotropic receptors, belong to the rhodopsin-type superfamily of receptors but the groups vary in rate of molecular evolution. Opsins are more conserved between species, although gene duplications exist (see Sison-Mangus et al. 2008; Pohl et al. 2009; McCulloch et al. 2017), while olfactory, gustatory and ionotropic receptors have duplicated extensively resulting in large gene families with a lot of copy number variation (Raible et al. 2006; van Schooten et al. 2016). Since these receptors have similar mechanisms of activation and similar functions in sensory perception, it is possible that the hydrophobic molecules with which they interact can be transported by proteins that are also similar to each other. In the cotton bollworm *H. armigera* four chemosensory proteins are expressed in both the eyes and proboscis; these proteins bind β-carotene and retinol (Zhu et al. 2016). That study demonstrates that proteins belonging to a family that responds to chemicals can have modified functions to have a role as a carrier for dietary carotenoids and visual processing in insects. Likewise, it is possible that *Hme CTD11* and *12,* upregulated in antennae, have functions in mediating olfaction through subfunctionalization.

Hme CTD31 candidate chromophore transporter

Hme CTD31 is a candidate to explore for functions in visual pigment transport due to its upregulation in heads. However, head libraries were generated using whole head mRNA, so we used reverse transcription PCR (RT-PCR) to dissect whether *Hme CTD31* was expressed in the eye, brain, or both. We used the 18S rRNA gene as a positive control for normalized mRNA presence. We also used *Hme CTD12* and antennae tissue to validate TPM expression patterns. We expected to see *Hme CTD31* expressed in the eye and brain but not in the antennae, and *Hme CTD12* only expressed in the antennae. *Hme CTD12* was only amplified in the antennae as expected (Figure 2.4). However, RT-PCR showed that *Hme CTD31* was only expressed in male and female eyes and not in the brain or the antennae (Figure 2.4). Additional support for *Hme CTD31* having a potential role in butterfly vision came from exploring the expression of CRAL-TRIO domain genes in heads of a different butterfly species, *B. anynana* (Macias-Muñoz et al. 2016) (accession numbers E-MTAB-3887 and doi:10.5061/dryad.f98s6). We found that the *B. anynana* ortholog of *Hme CTD31* is the most highly expressed CRAL-TRIO domain gene in *Bicyclus* butterfly heads (supplementary fig. S8) further supporting that expression of this gene is important in the compound eye across butterfly species.

To localize where the Hme CTD31 protein is expressed in the *H. melpomene* eye, we designed an antibody against a unique peptide to perform immunohistochemistry. Our protein of interest has a predicted weight of ~35 kilodaltons, and an immunoblot of proteins extracted from whole head tissue using this antibody indicates it binds to a protein of the expected size (Figure 2.5A). We saw another band below 35 kDa and that maybe the same protein but running through the gel differently due to phosphorylation of specific residues in the protein. Hme CTD31 has sites that are potentially phosphorylated

with a probability score higher than 0.75 at sites 7, 74, 109, 127, 175, 233, 275, and 28 (Blom et al. 1999) .

Next, to identify the cellular localization of the protein we examined longitudinal and transverse sections of the butterfly compound eye (Figure 2.6A). Each *Heliconius* ommatidium consists of a cornea, crystalline cone, and 9 photoreceptor cells with a fused rhabdom and a tiered cell body arrangement. Primary pigment cells surround the crystalline cone and secondary pigment cells surround the photoreceptor cells. Brightfield images of a longitudinal section of the compound eye showed that there is pigment at the top of the ommatidia, around or within each ommatidium for its entire length, and below the basement membrane in tracheal cells (fig.6B). A transverse image showed that the ommatidia are surrounded by 8 tracheoles which have pigment along the tracheal walls (Figure 2.6D).

We used polyclonal antibodies against Hme CTD31 and the long wavelength opsin (LWRh) to visualize where Hme CTD31 was expressed in relation to photoreceptor cells (McCulloch et al. 2016). We found that Hme CTD31 is found in the primary pigment cells, secondary pigment cells and tracheal cells (Figure 2.6C). The tracheal cells project tracheoles up and around the ommatidia, and these structures autofluoresce under blue light (488 nm laser) due to the presence of chitin (Figure 2.6E‴) (Iwata et al. 2014). Hme CTD31 is also expressed in the cell bodies surrounding the tracheole walls (Figure 2.6E). Hme CTD31 immunohistochemical results were similar to those of a retinol binding protein in the family Papilionidae, *Papilio* retinol binding protein (RBP). *Papilio* RBP binds retinol and was found to be expressed in primary pigment cells, secondary pigment cells and tracheal cells (Wakakuwa et al. 2004). However, Hme CTD31 is expressed in the lower two-

thirds of the ommatidia, rather than along the entire length, while *Papilio* RBP is found in the entire length of the ommatidia. The difference in where *Papilio* RBP and Hme CTD31 are located in *Papilio* and *Heliconius* respectively might be due to the difference in ommatidium morphology. *Papilio* RBP also does not belong to the CRAL-TRIO domain gene family. However, an ortholog of the gene encoding *Papilio* RBP in *H. melpomene* (Hme comp30064) was upregulated in heads relative to other tissue types (Table S2.6-8; supplementary fig. S9). It is possible that Hme CTD31 and *Papilio* RBP are both necessary to transport the retinal molecule in different configurations as in vertebrates (McBee et al. 2001). The study characterizing *pinta* suggested there might be additional proteins in the primary pigment cells that are required for biosynthesis of the chromophore (Wang & Montell 2005).

It is also possible that Hme CTD31 functions in binding filtering pigments. From the RT-PCR and immunohistochemistry alone, we cannot confirm to what molecule Hme CTD31 binds but its upregulation in heads and localization in the ommatidia suggest that this protein has a role in butterfly vision. To confirm whether Hme CTD31 binds a chromophore, proteins from butterfly heads were run on a native gel and examined under UV light (Figure 2.5B). In the swallowtail butterfly, *Papilio* RBP bound to the chromophore fluoresces under UV light (Wakakuwa et al. 2004). We found 3 fluorescing bands which were cut and sequenced using mass spectrometry. Our candidate protein Hme CTD 31 is one of the top 20 proteins matching peptide fragment fluorescing upper (consisting of insoluble material), middle and lower bands as detected by mass spectrometry (Table 2.3). This evidence further supports our hypothesis that Hme CTD31, a CRAL-TRIO domain containing protein, is binding the chromophore molecule in butterflies. Hme CTD31 likely

transports the vitamin-A derived chromophore molecule similar to vertebrate CRALBP and *Drosophila* PINTA. *Drosophila cralbp* and *pinta* both belong to the CRAL-TRIO and SEC14 superfamilies yet PINTA is the one shown experimentally to bind retinal (Wang & Montell 2005). Similarly, RALBP also belongs to the SEC14 superfamily and also functions in retinal binding in cephalopods (Ozaki et al. 1994; Speiser et al. 2014). These previous findings and our results suggest a conserved role for at least some of the CRAL-TRIO domain proteins, even if the specific function in this pathway is undertaken by non-orthologous members of the expanded gene family.

To summarize, we investigated a large gene family whose function in insects is only known for one gene in one species: *pinta* transports the chromophore molecule in *Drosophila* and is necessary for phototransduction. While other members of the CRAL-TRIO domain gene family have undergone an expansion, we found no ortholog of the *pinta* gene in Lepidoptera. In *H. melpomene*, we found an expansion of genes in close proximity suggesting that CRAL-TRIO domain genes are evolving by tandem duplications. We also found copy number variation of CRAL-TRIO domain genes between individuals. While the function of these genes is not known, we hypothesized that one or more of these genes could have a role in vision similar to *pinta* and we were able to identify one candidate gene upregulated in *H. melpomene* heads and two other genes upregulated in antennae. This gene, *Hme CTD31*, was found in eye mRNA and its protein product was localized to secondary and primary pigment cells and to a protein gel band that fluoresces under UV light. Interestingly, *Hme CTD31* is a single copy gene across the 18 resequenced genomes we investigated, suggesting a critical function. We have thus identified a CRAL-TRIO

domain containing gene that likely encodes a chromophore binding protein in butterflies, a paralogous member of the *pinta* gene family that is rapidly evolving in butterflies.

ACKNOWLEDGMENTS

We thank Ana Catalán, Gilbert Smith, Ali Mortazavi, Kevin Thornton, JJ Emerson, and Francisco Ayala for advice on data handling, and Jennifer Briner for comments on manuscript. We also thank James Baldwin-Brown and Mahul Chakraborty for bioinformatic assistance, and Junsung Im and Paul Gershon for technical assistance with the mass spectrometry. This work was supported in part from National Science Foundation grants (IOS-1257627 and IOS-1656260 to ADB) and a National Science Foundation Graduate Research Fellowship to AMM.

REFERENCES

Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 12:363–376. doi: 10.1038/nrg2958.

Baumgarten A, Cannon S, Spangler R, May G. 2003. Genome-level evolution of resistance genes in *Arabidopsis thaliana*. Genetics. 165:309–319.

Blom N, Gammeltoft S, Brunak S. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294:1351–62. doi: 10.1006/jmbi.1999.3310.

Bomar JM et al. 2003. Mutations in a novel gene encoding a CRAL-TRIO domain cause human Cayman ataxia and ataxia/dystonia in the jittery mouse. Nat. Genet. 35:264–269. doi: 10.1038/ng1255.

Briscoe AD et al. 2013. Female behaviour drives expression and evolution of gustatory receptors in butterflies. PLoS Genet. 9:e1003620. doi: 10.1371/journal.pgen.1003620.

Camacho C et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421. doi: Artn 421\nDoi 10.1186/1471-2105-10-421.

Cannon SB, Mitra A, Baumgarten A, Young ND, May G. 2004. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana.* BMC Plant Biol. 4:10. doi: 10.1186/1471-2229-4-10.

Catalán A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J. 2016. An indel polymorphism in the *MtnA 3'* untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*. PLoS Genet. 12:1–24. doi: 10.1371/journal.pgen.1005987.

Conrad DF et al. 2010. Origins and functional impact of copy number variation in the human genome. Nature. 464:704–712. doi: 10.1038/nature08516.

Cooper GM, Nickerson D a, Eichler EE. 2007. Mutational and selective effects on copy-number variants in the human genome. Nat. Genet. 39:S22-9. doi: 10.1038/ng2054.

Dabney A, Storey JD. 2013. qvalue: Q-value estimation for false discovery rate control.

Davey JW et al. 2016. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 Million years of butterfly evolution. G3. 6:695–708.

Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. 2006. The evolution of mammalian gene families. PLoS One. 1:e85. doi: 10.1371/journal.pone.0000085.

Dopman EB, Hartl DL. 2007. A portrait of copy-number polymorphism in *Drosophila* melanogaster. Proc Nat Acad Sci USA. 104:19920–5. doi: 10.1073/pnas.0709888104.

Duvaux L et al. 2014. Dynamics of copy number variation in host races of the pea aphid. Mol. Biol. Evol. 32:63–80. doi: 10.1093/molbev/msu266.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.

Emerson JJ, Cardoso-Moreira M, Borevitz JO, Long M. 2008. Natural selection shapes genome-wide patterns of copy-number polymorphism in *Drosophila melanogaster*. Science. 320:1629–1631. doi: 10.1126/science.1158078.

Frentiu FD et al. 2007. Adaptive evolution of color vision as seen through the eyes of butterflies. Proc. Natl. Acad. Sci. U. S. A. 104 Suppl:8634–8640. doi: 10.1073/pnas.0701447104.

Grabherr MG et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–52. doi: 10.1038/nbt.1883.

Haas BJ et al. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512. doi: 10.1038/nprot.2013.084.

Henikoff S. 1997. Gene families: The taxonomy of protein paralogs and chimeras. Science. 278:609–614. doi: 10.1126/science.278.5338.609.

Henikoff S, Henikoff JG. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. 89:10915–10919. doi: 10.1073/pnas.89.22.10915.

Hersh BM, Carroll SB. 2005. Direct regulation of knot gene expression by Ultrabithorax and the evolution of cis-regulatory elements in *Drosophila*. Development. 132:1567–1577. doi: 10.1242/dev.01737.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755. doi: 10.1093/bioinformatics/17.8.754.

Iwata M, Ohno Y, Otaki JM. 2014. Real-time in vivo imaging of butterfly wing development: Revealing the cellular dynamics of the pupal wing tissue. PLoS One. 9:e89500. doi: 10.1371/journal.pone.0089500.

Jelesko JG, Harper R, Furuya M, Gruissem W. 1999. Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. U. S. A. 96:10302–10307. doi: 10.1073/pnas.96.18.10302.

Kanost MR et al. 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. Insect Biochem. Mol. Biol. 76:118–147. doi: 10.1016/j.ibmb.2016.07.005.

Koressaar T, Remm M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics. 23:1289–1291. doi: 10.1093/bioinformatics/btm091.

Kozak KM et al. 2015. Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. Syst. Biol. 64:505–524. doi: 10.1093/sysbio/syv007.

Lespinet O, Wolf YI, Koonin E V., Aravind L. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12:1048–1059. doi: 10.1101/gr.174302.

Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44:W242–W245. doi: 10.1093/nar/gkw290.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323. doi: 10.1186/1471-2105-12-323.

Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079. doi: 10.1093/bioinformatics/btp352.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760. doi: 10.1093/bioinformatics/btp324.

von Lintig J, Kiser PD, Golczak M, Palczewski K. 2010. The biochemical and structural basis for trans-to-cis isomerization of retinoids in the chemistry of vision. Trends Biochem. Sci. 35:400–410. doi: 10.1016/j.tibs.2010.01.005.

Long M, Betrán E, Thornton K, Wang W. 2003. The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. 4:865–875. doi: 10.1038/nrg1204.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science. 290:1151–1155. doi: 10.1126/science.290.5494.1151.

Macias-Muñoz A, Smith G, Monteiro A, Briscoe AD. 2016. Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic. Mol. Biol. Evol. 33:79–92. doi: 10.1093/molbev/msv197.

Martin SH et al. 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23:1817–1828. doi: 10.1101/gr.159426.113.

Maw MA et al. 1997. Mutation of the gene encoding cellular retinaldehyde-binding protein in autosomal recessive retinitis pigmentosa. Nat. Genet. 15:198–200.

McBee JK, Palczewski K, Baehr W, Pepperberg DR. 2001. Confronting complexity: The interlink of phototransduction and retinoid metabolism in the vertebrate retina. Prog. Retin. Eye Res. 20:469–529. doi: 10.1016/S1350-9462(01)00002-7.

Mccarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. 40:4288–4297. doi: 10.1093/nar/gks042.

McCulloch KJ et al. 2017. Sexual dimorphism and retinal mosaic diversification following the evolution of a violet receptor in butterflies. 1–14. doi: 10.1093/molbev/msx163.

McCulloch KJ, Osorio D, Briscoe AD. 2016. Sexual dimorphism in the compound eye of *Heliconius erato*: a nymphalid butterfly with at least five spectral classes of photoreceptor. J. Exp. Biol. 219:2377–2387. doi: 10.1242/jeb.136523.

Min KC, Kovall RA, Hendrickson WA. 2003. Crystal structure of human alpha-tocopherol transfer protein bound to its ligand: implications for ataxia with vitamin E deficiency. Proc. Natl. Acad. Sci. U. S. A. 100:14713–14718. doi: 10.1073/pnas.2136684100.

Nadeau NJ et al. 2016. The gene *cortex* controls mimicry and crypsis in butterflies and moths. Nature. 534:106–110. doi: 10.1038/nature17961.

Nei M, Niimura Y, Nozawa M. 2008. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat. Rev. Genet. 9:951–963. doi: 10.1038/nrg2480.

Nozawa M, Nei M. 2008. Genomic drift and copy number variation of chemosensory receptor genes in humans and mice. Cytogenet. Genome Res. 123:263–269. doi: 10.1159/000184716.

Ohno S. 1970. Evolution by gene duplication. Springer-Verlag: New York.

Ozaki K et al. 1994. Molecular characterization and functional expression of squid retinal-binding protein. J. Biol. Chem. 269:3838–3845.

Panagabko C et al. 2003. Ligand specificity in the CRAL-TRIO protein family. Biochemistry. 42:6467–6474. doi: 10.1021/bi034086v.

Pinharanda A, Martin SH, Barker SL, Davey JW, Jiggins CD. 2016. Data from: The comparative landscape of duplications in *Heliconius melpomene* and *Heliconius cydno*. Dryad Digit. Repos.

Pinharanda A, Martin SH, Barker SL, Davey JW, Jiggins CD. 2017. The comparative landscape of duplications in *Heliconius melpomene* and *Heliconius cydno*. Heredity. 118:78–87. doi: 10.1038/hdy.2016.107.

Pohl N, Sison-Mangus MP, Yee EN, Liswi SW, Briscoe AD. 2009. Impact of duplicate gene copies on phylogenetic analysis and divergence time estimates in butterflies. BMC Evol. Biol. 9:99. doi: 10.1186/1471-2148-9-99.

Raible F et al. 2006. Opsins and clusters of sensory G-protein-coupled receptors in the sea urchin genome. Dev. Biol. 300:461–475. doi: 10.1016/j.ydbio.2006.08.070.

Redon R et al. 2006. Global variation in copy number in the human genome. Nature. 444:444–54. doi: 10.1038/nature05329.

Robinson M, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25. doi: 10.1186/gb-2010-11-3-r25.

Robinson MD, Mccarthy DJ, Smyth GK. 2010. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. 26:139–140. doi: 10.1093/bioinformatics/btp616.

Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics. 23:2881–2887. doi: 10.1093/bioinformatics/btm453.

Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics. 9:321–332. doi: 10.1093/biostatistics/kxm030.

Ronquist F, Huelsenbeck J, Teslenko M. 2011. MrBayes Version 3.2 Manual: Tutorials and Model Summaries. Man. MrBayes. 1–103.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574. doi: 10.1093/bioinformatics/btg180.

Schmidt JM et al. 2010. Copy number variation and transposable elements feature in recent, ongoing adaptation at the *Cyp6g1* locus. PLoS Genet. 6:1–11. doi: 10.1371/journal.pgen.1000998.

van Schooten B, Jiggins CD, Briscoe AD, Papa R. 2016. Genome-wide analysis of ionotropic receptors provides insight into their evolution in *Heliconius* butterflies. BMC Genomics. 17:254. doi: 10.1186/s12864-016-2572-y.

Sharp AJ et al. 2005. Segmental duplications and copy-number variation in the human genome. Am. J. Hum. Genet. 77:78–88. doi: 10.1086/431652.

Sigrist CJA et al. 2013. New and continuing developments at PROSITE. Nucleic Acids Res. 41:344–347. doi: 10.1093/nar/gks1067.

Sison-Mangus MP, Briscoe AD, Zaccardi G, Knuttel H, Kelber A. 2008. The lycaenid butterfly *Polyommatus icarus* uses a duplicated blue opsin to see green. J. Exp. Biol. 211:361–369. doi: 10.1242/jeb.012617.

Smith G, Briscoe AD. 2015. Molecular evolution and expression of the CRAL_TRIO protein family in insects. Insect Biochem. Mol. Biol. 62:168–173. doi: 10.1016/j.ibmb.2015.02.003.

Smith G, Macias-Muñoz A, Briscoe AD. 2016. Gene duplication and gene expression changes play a role in the evolution of candidate pollen feeding genes in *Heliconius* butterflies. Genome Biol. Evol. 8:2581–96. doi: 10.1093/gbe/evw180.

Speiser DI et al. 2014. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. BMC Bioinformatics. 15:350. doi: 10.1186/s12859-014-0350-x.

Stanke M, Morgenstern B. 2005. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 33:465–467. doi: 10.1093/nar/gki458.

Stern DL. 2000. Evolutionary developmental biology and the problem of variation. Evolution. 54:1079–1091. doi: 10.1111/j.0014-3820.2000.tb00544.x.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100:9440–5. doi: 10.1073/pnas.1530509100.

Stranger BE et al. 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 315:848–853.

Tarazona S, García-Alcalde F, Dopazo J, Alberto F, Conesa A. 2011. Differential expression in RNA-seq: A matter of depth. Genome Res. 2213–2223. doi: 10.1101/gr.124321.111.Freely.

Troemel ER, Chou JH, Dwyer ND, Colbert HA, Bargmann CI. 1995. Divergent seven transmembrane receptors are candidate chemosensory receptors in *C. elegans*. Cell. 83:207–218. doi: 10.1016/0092-8674(95)90162-0.

Untergasser A et al. 2012. Primer3-new capabilities and interfaces. Nucleic Acids Res. 40:e115. doi: 10.1093/nar/gks596.

Wakakuwa M, Ozaki K, Arikawa K. 2004. Immunohistochemical localization of *Papilio RBP* in the eye of butterflies. J. Exp. Biol. 207:1479–1486. doi: 10.1242/jeb.00913.

Wang T, Montell C. 2005. Rhodopsin formation in *Drosophila* is dependent on the PINTA retinoid-binding protein. J. Neurosci. 25:5187–94.

Wu Z et al. 2006. CRALBP ligand and protein interactions. Adv. Exp. Med. Biol. 572:477–483.

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 25:2865–2871. doi: 10.1093/bioinformatics/btp394.

Zhang J. 2003. Evolution by gene duplication: An update. Trends Ecol. Evol. 18:292–298. doi: 10.1016/S0169-5347(03)00033-8.

Zhu J et al. 2016. Conserved chemosensory proteins in the proboscis and eyes of Lepidoptera. Int. J. Biol. Sci. 12:1394–1404. doi: 10.7150/ijbs.16517.

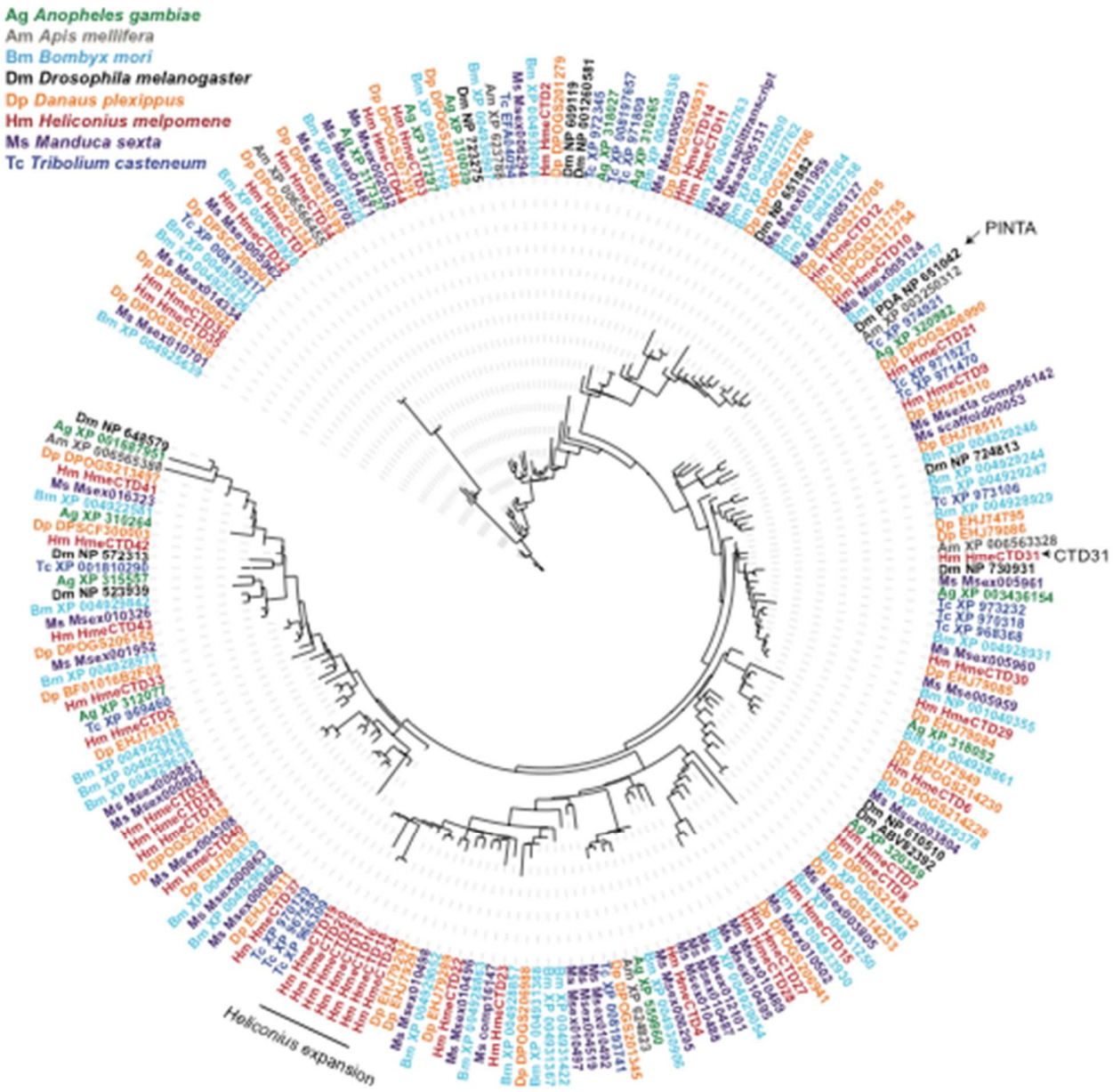


Figure 2.1. Bayesian phylogeny of insect CRAL-TRIO domain proteins. Phylogeny includes sequences from *Anopheles gambiae* (green), *Apis mellifera* (grey), *Bombyx mori* (light blue), *Drosophila melanogaster* (black), *Danaus plexippus* (orange), *Heliconius melpomene* (red), *Manduca sexta* (purple) and *Tribolium casteneum* (dark blue). The Bayesian tree was found using MrBayes with a BLOSUM62 model of amino acid substitution. The *Heliconius* expansion as well as *Drosophila pinta* and *Heliconius Hme CTD31* are indicated on the phylogeny with black lines and arrows.

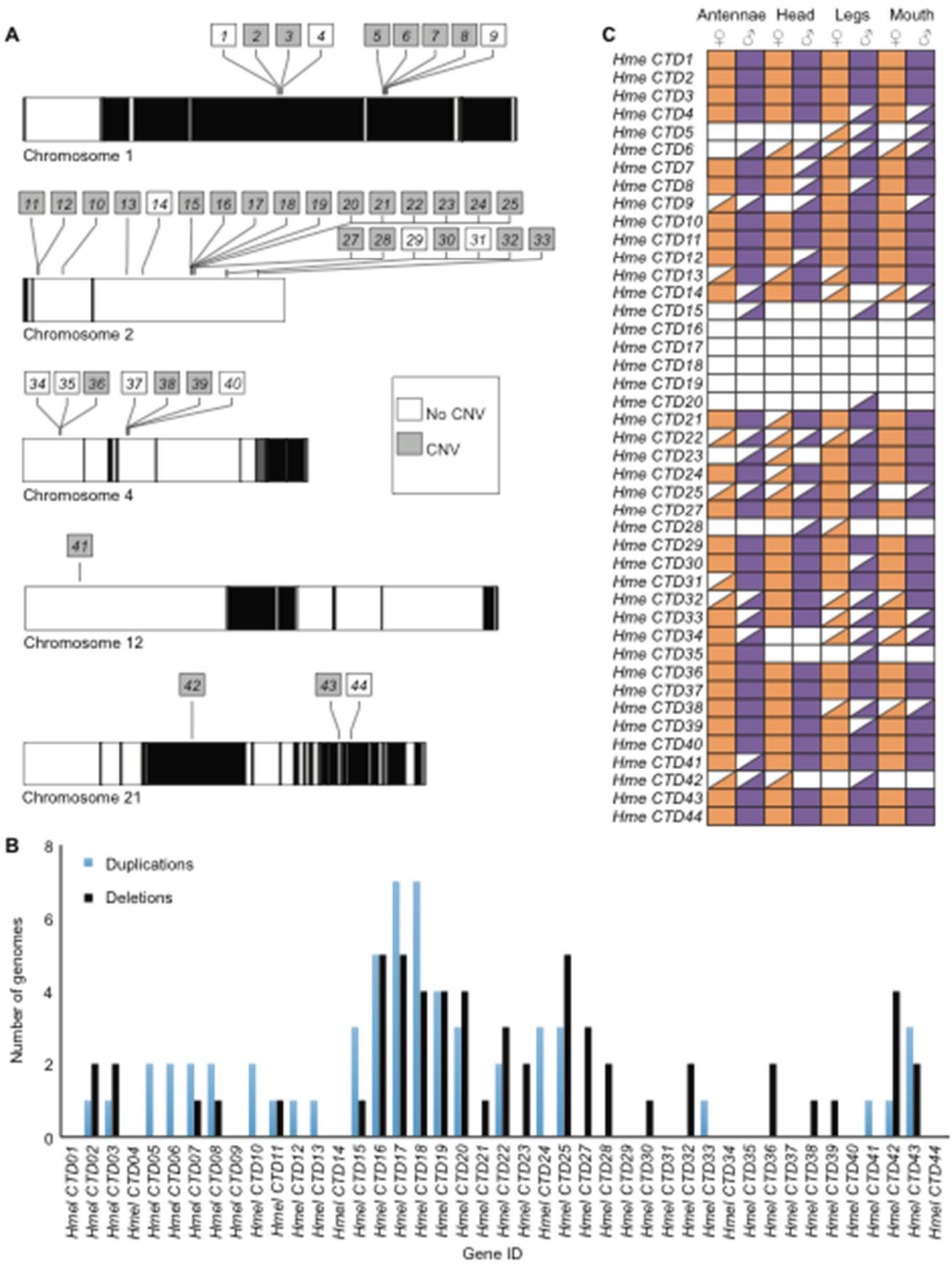


Figure 2.2. CRAL-TRIO domain gene location and patterns of mRNA presence or absence. (A) CRAL-TRIO domain genes are located on 5 chromosomes, many in tandem. Alternating black and white chromosomal regions represent scaffolds. Shaded squares

represent genes with copy number variation, duplicated and/or deleted in at least one of 18 resequenced genomes. (B) Number of genomes in which CRAL-TRIO domain genes are deleted (black) or duplicated (blue) in 18 *H. melpomene* resequenced genomes. (C) mRNA presence patterns of CRAL-TRIO domain genes in *H. melpomene* male and female antennae, head, legs and mouth parts. Filled square represents complete presence (> 1 TPM for all replicates), half-filled square represents partial presence (> 1 TPM for at least one replicate but not all four), and no fill represents lack of transcript mRNA.

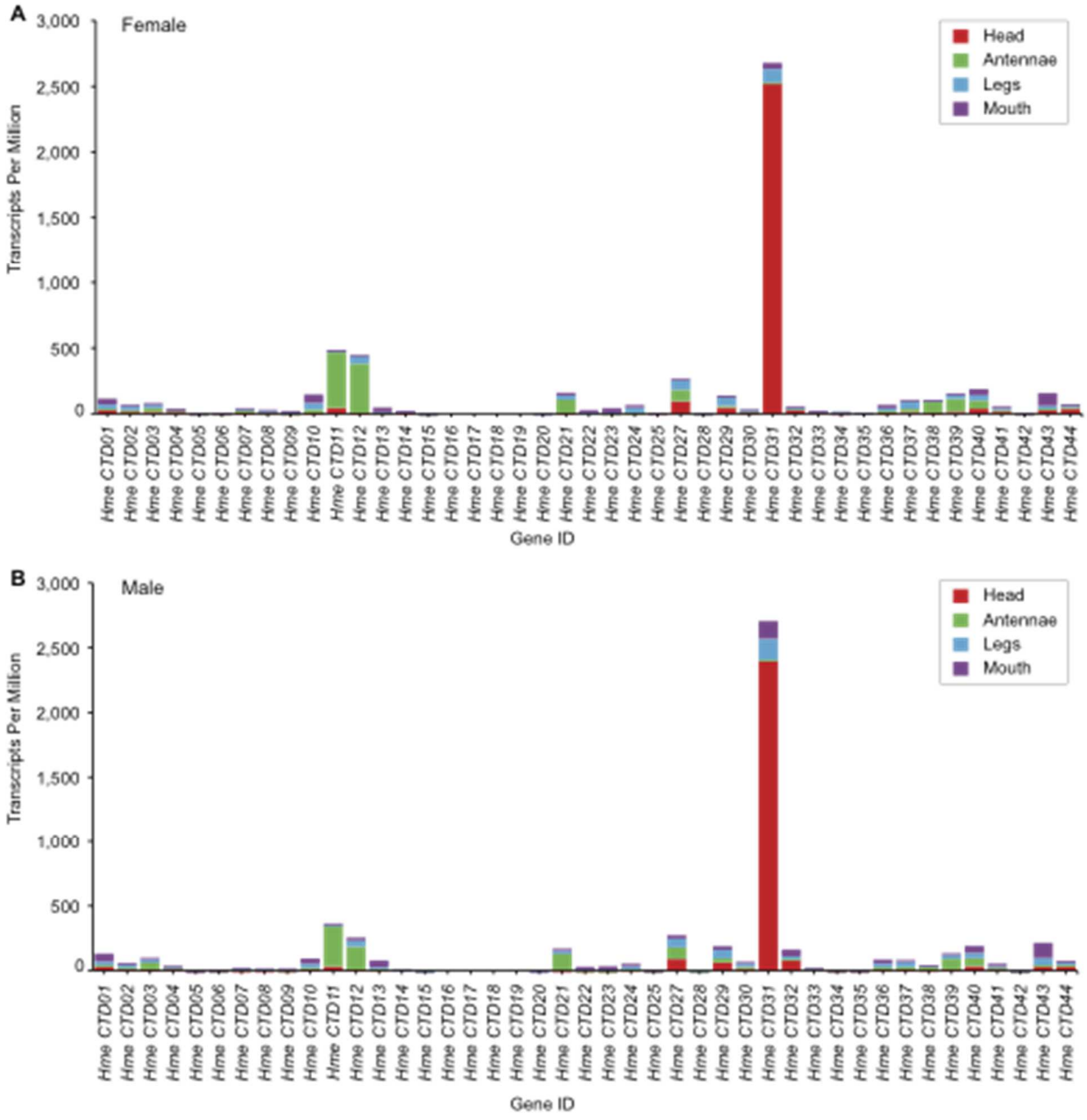


Figure 2.3. Expression of CRAL-TRIO domain genes. (A) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in female antennae, head, legs and mouth parts. (B) TPM of CRAL-TRIO domain genes in male antennae, head, legs and mouth parts.

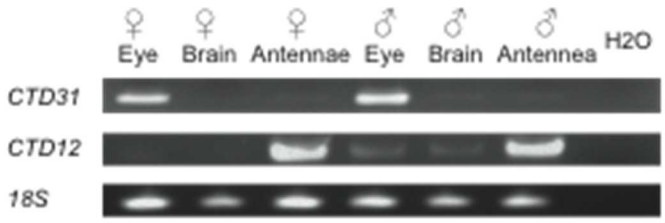


Figure 2.4. *Hme CTD31* RT-PCR. RT-PCR of *Hme CTD31*, *Hme CTD12* and *18S* in female and male eye, brain, and antennae.

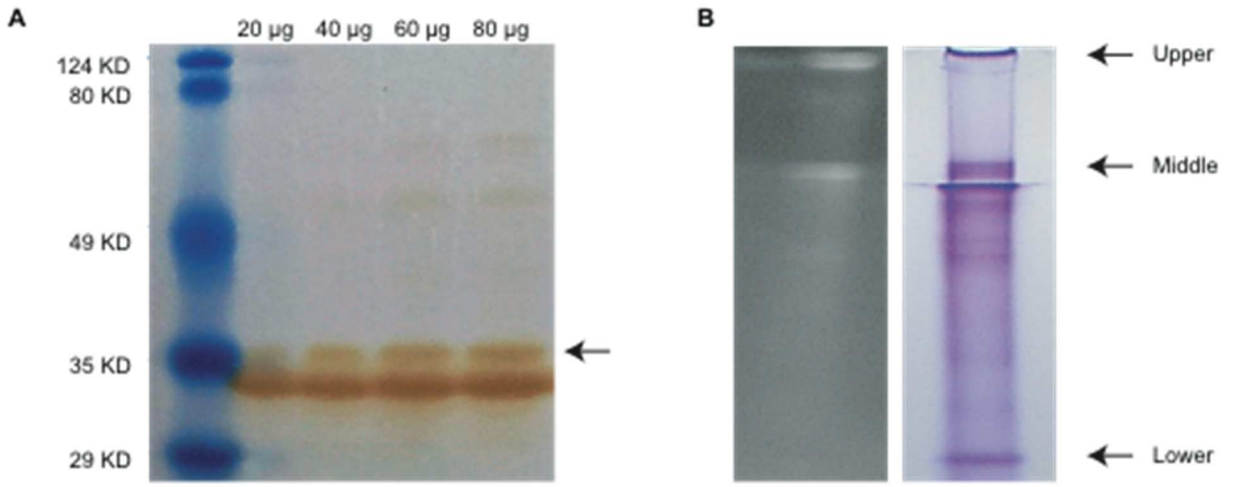


Figure 2.5. *Hme CTD31* Western Blot. (A) Western blot using head tissue and *Hme CTD31* antibody performed by Zyagen (San Diego CA) arrow indicates expected band. (B) Butterfly head protein run on native gel shows 3 bands that fluoresce under UV light. Arrows indicate the location of the upper, middle, and lower bands which were cut out and sequenced using mass spectrometry.

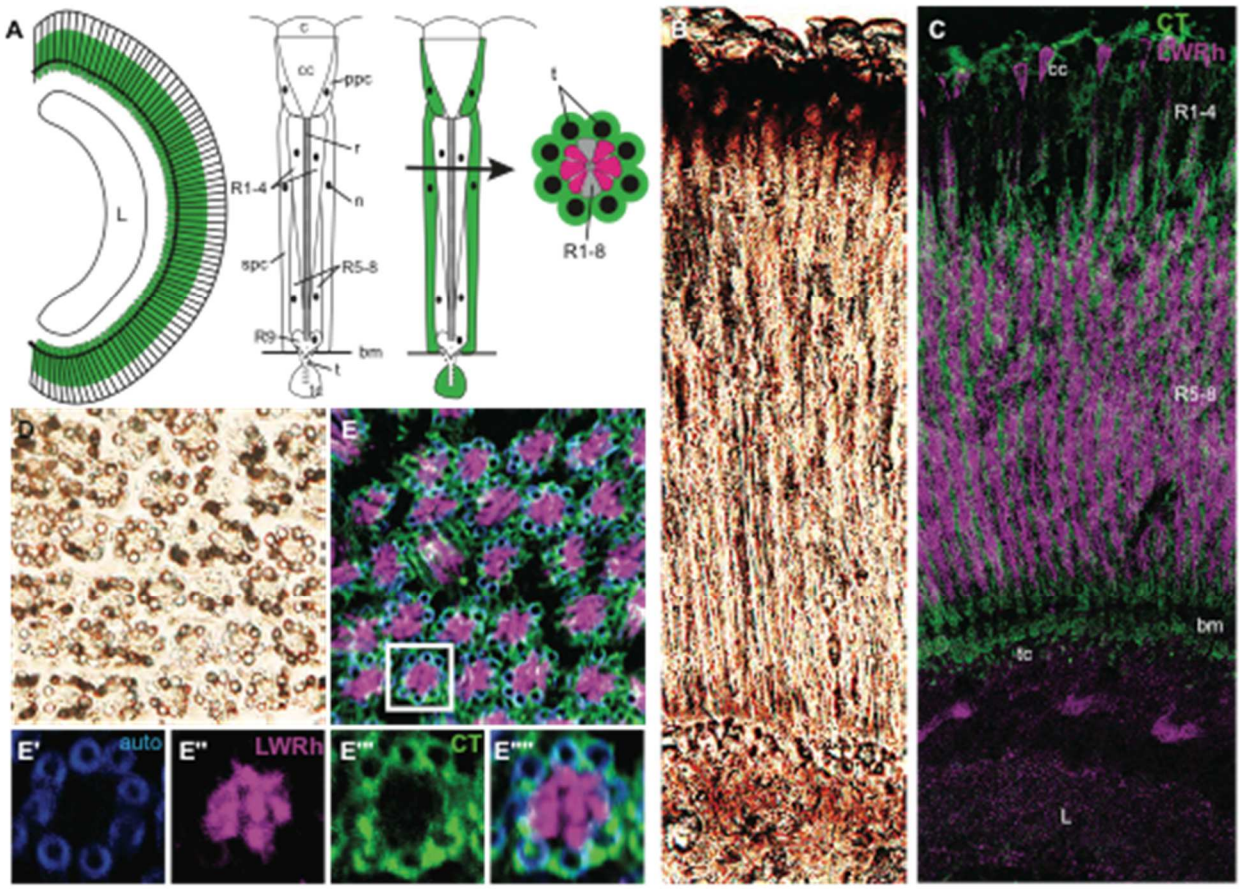


Figure 2.6. Immunohistochemistry of Hme CTD31 in *H. melpomene* eye and optic lobe. (A) Drawing of a longitudinal view of a compound eye and lamina, and longitudinal and transverse sections of a single ommatidium. Green highlights where we find Hme CTD31 expression; L, lamina; c, cornea; cc, crystalline cone; ppc, primary pigment cells; r, rhabdom; R1-9 conventional Lep. numbering of photoreceptor cells; n, cell nucleus; spc, secondary pigment cells; bm, basement membrane; t, trachea; and tc, tracheal cell. (B) Brightfield longitudinal section showing pigments in the *H. melpomene* retina. (C) Longitudinal section with Hme CTD31 and LW opsin staining; Hme CTD31 is in green and LW opsin is in magenta. (D) Brightfield image of a transverse section of a butterfly eye, pigment is seen in the structures surrounding the ommatidia. (E) Transverse view of a butterfly eye stained for LW and Hme CTD31. (E') autofluorescence showing tracheoles surrounding an individual ommatidium. (E'') LW opsin staining showing where the LW photoreceptor cells are. (E''') CTD31 staining showing where the CRAL-TRIO domain protein Hme CTD31 is expressed. (E'''') merged image of LWRh, CTD31 and autofluorescence.

Table 2.1. Summary of differentially expressed (DE) and upregulated contigs.

	qvalue	Bonferroni	Upregulated in Heads*
Head vs. Antennae	4,868	1,173	561
Head vs. Legs	6,108	1,472	928
Head vs. Mouth	6,176	1,486	914

*These contigs are upregulated in heads using a Bonferroni FDR correction

Table 2.2. Expression patterns of CRAL-TRIO domain contigs.

	qvalue			Bonferroni		
Gene ID	**H vs. A**	**H vs. L**	**H vs. M**	**H vs. A**	**H vs. L**	**H vs. M**
Hme CTD1	Not DE	Not DE	Down	Not DE	Not DE	Not DE
Hme CTD2	Not DE	Down	Down	Not DE	Not DE	Not DE
Hme CTD3	Not DE	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD4	Down	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD6	Not DE	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD8	Not DE	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD9	Down	Down	Down	Not DE	Not DE	Not DE
Hme CTD10	Down	Down	Down	Down	Down	Not DE
Hme CTD11	Down	Not DE	Not DE	Down	Not DE	Not DE
Hme CTD12	Down	Down	Down	Down	Down	Not DE
Hme CTD13	Down	Down	Down	Down	Not DE	Down
Hme CTD14	Down	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD15	Not DE	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD21	Down	Down	Down	Down	Down	Down
Hme CTD22	**Up**	**Up**	Down	Not DE	Not DE	Not DE
Hme CTD29	Not DE	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD30	Down	Down	Not DE	Not DE	Not DE	Not DE
Hme CTD31	**Up**	**Up**	**Up**	**Up**	Not DE	Not DE
Hme CTD32	Not DE	Not DE	Not DE	Not DE	Not DE	Not DE
Hme CTD34	Down	Down	Down	Down	Down	Not DE
Hme CTD35	Down	Not DE	Not DE	Down	Not DE	Not DE
Hme CTD36	Down	Down	Down	Down	Down	Down
Hme CTD37	Down	Down	Down	Down	Down	Not DE
Hme CTD38	Down	Not DE	Not DE	Down	Not DE	Not DE
Hme CTD39	Down	Down	Not DE	Not DE	Not DE	Not DE
Hme CTD40	Down	Not DE	Down	Not DE	Not DE	Not DE
Hme CTD41	Down	Down	Down	Not DE	Not DE	Not DE
Hme CTD43	Not DE	Not DE	Down	Not DE	Not DE	Down

Not DE - Not Differentially Expressed
Up - Upregulated in heads
Down - Downregulated in heads

Table 2.3. Top 20 proteins from upper, middle and lower bands detected by mass spectrometry sorted by upper band Mascot score.

		Upper		Middle		Lower	
Accession	**Protein Family**	**Mascot Score**	**Peptide Matches**	**Mascot Score**	**Peptide Matches**	**Mascot Score**	**Peptide Matches**
comp33735_c0	Rfabg	5766	198	1699	67	45	2
comp31078_c1	Atpalpha	2204	62	1557	41	561	16
comp31397_c0	betaTub56D	1587	47	925	26	594	18
comp27767_c0	Vha68-2	1542	39	946	24	298	12
comp32095_c0	ATPsynbeta	1488	41	855	25	719	25
comp28890_c0	Gapdh2	1204	33	820	26	368	8
comp15204_c0	kdn	1187	31	787	24	660	17
comp27239_c0	CG1635	1051	29	692	18	127	2
comp31202_c0	alpha-Spec	998	38	1262	51	39	3
comp26414_c0	PyK	997	30	784	19	483	13
comp31948_c0	Pp2A-29B	963	20	475	17	584	19
comp29636_c0	**CG2663**	**947**	**43**	**514**	**24**	**494**	**24**
comp33018_c0	TER94	890	29	534	15	227	10
comp30615_c0	nrv3	867	19	786	21	139	4
comp14607_c0	Pgi	836	24	667	23	358	11
comp29963_c1	Gdh	833	29	1103	40	677	27
comp28746_c0	blw	822	27	588	19	57	3
comp25729_c0	Hsc70-4	817	32	2565	93	945	32
comp29025_c0	Mdh2	817	25	764	21	200	7
comp31520_c0	alphaTub84B	741	26	609	16	262	10
comp30064_c0	CG10476	517	12	271	5	309	7

Mascot protein scores greater than 67 are significant ($p<0.05$). Bold indicates comp29636_c0, which corresponds to CTD31. comp30064 corresponds to the *Papilio* RBP homolog in *H. melpomene*.

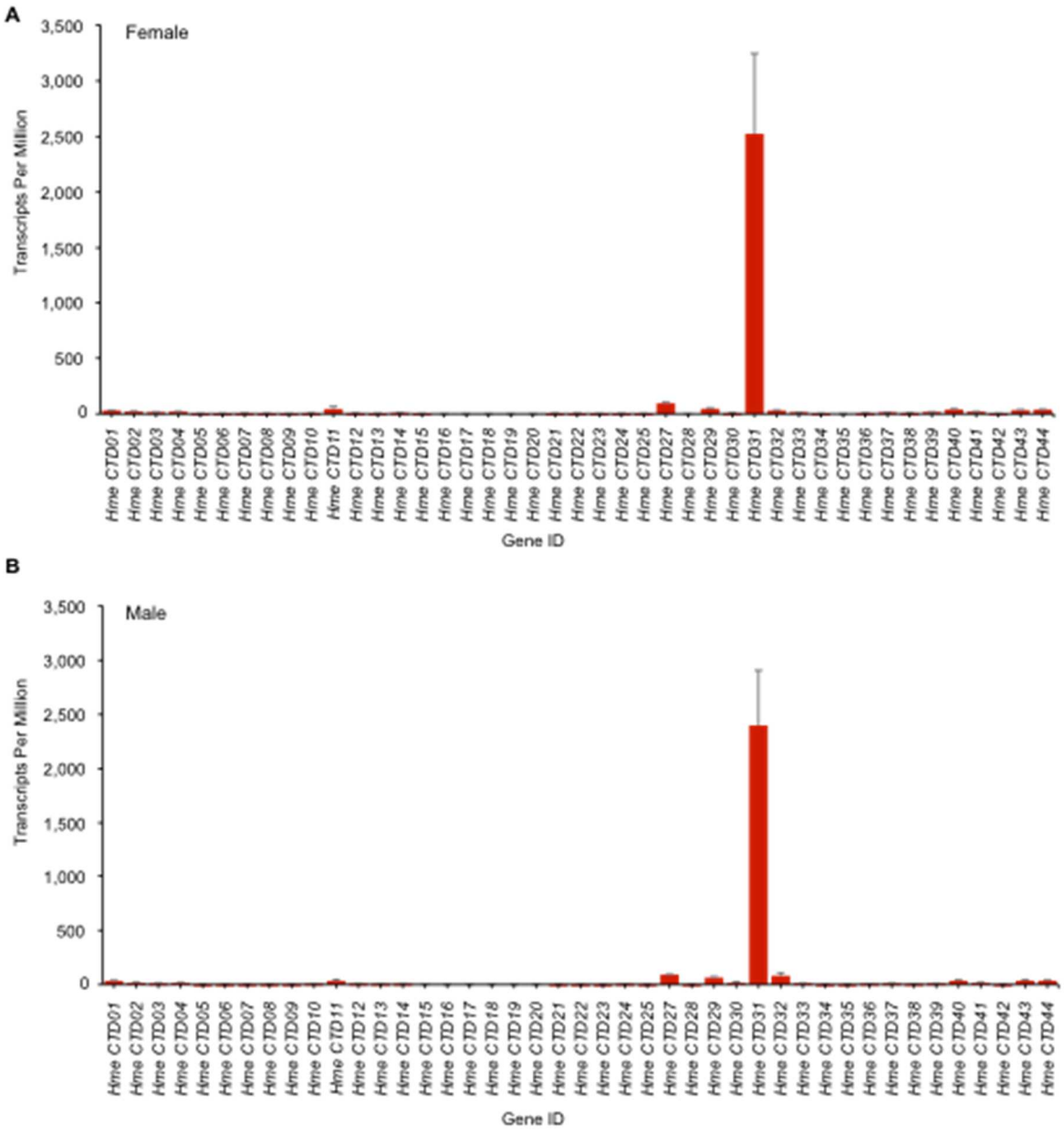


Figure S2.1. CRAL-TRIO domain gene expression in *H. melpomene* heads. (A) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* female head with standard error bars. (B) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* male head with standard error bars.

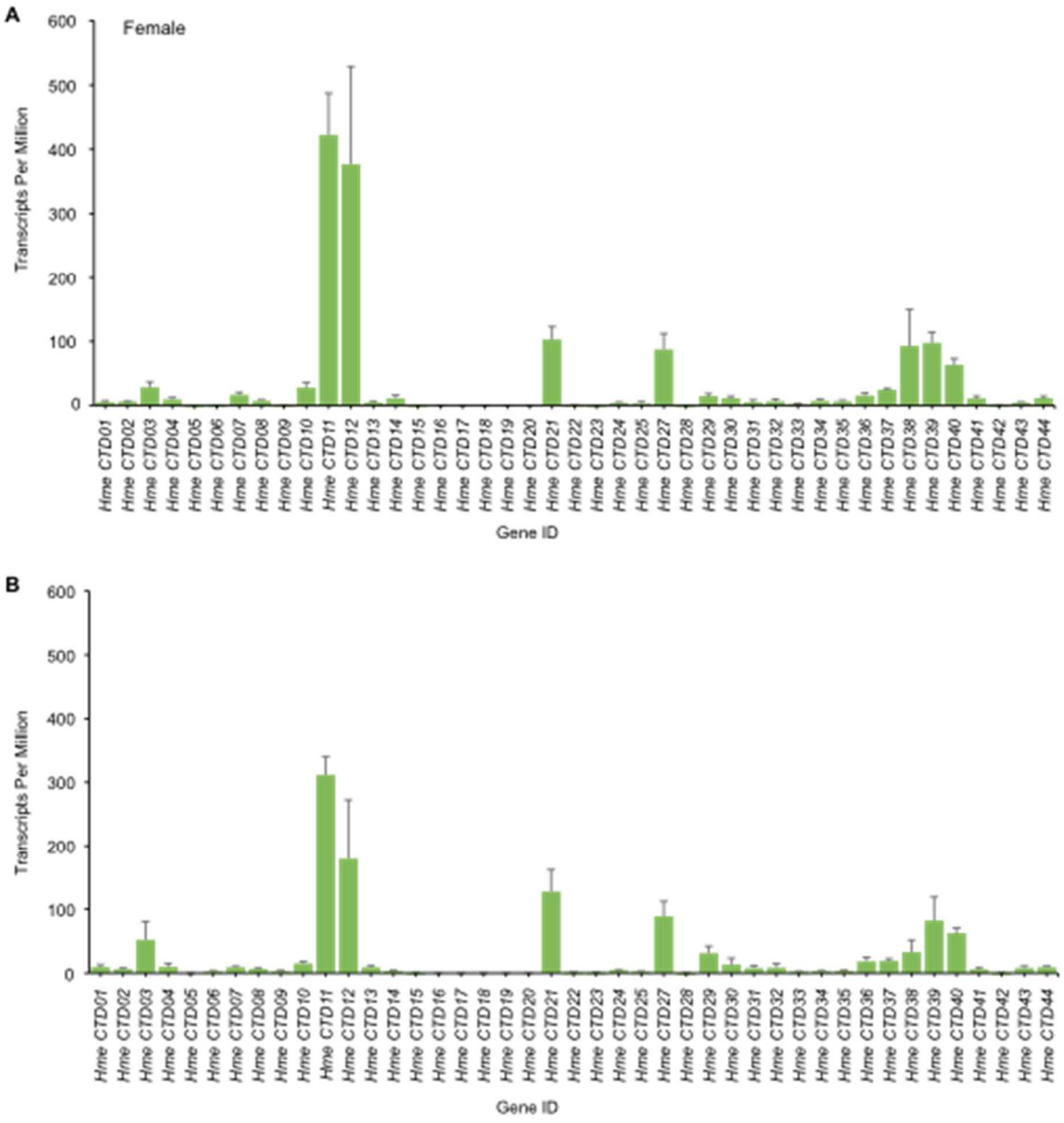


Figure S2.2. CRAL-TRIO domain gene expression in *H. melpomene* antennae. (A) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* female antennae with standard error bars. (B) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* male antennae with standard error bars.

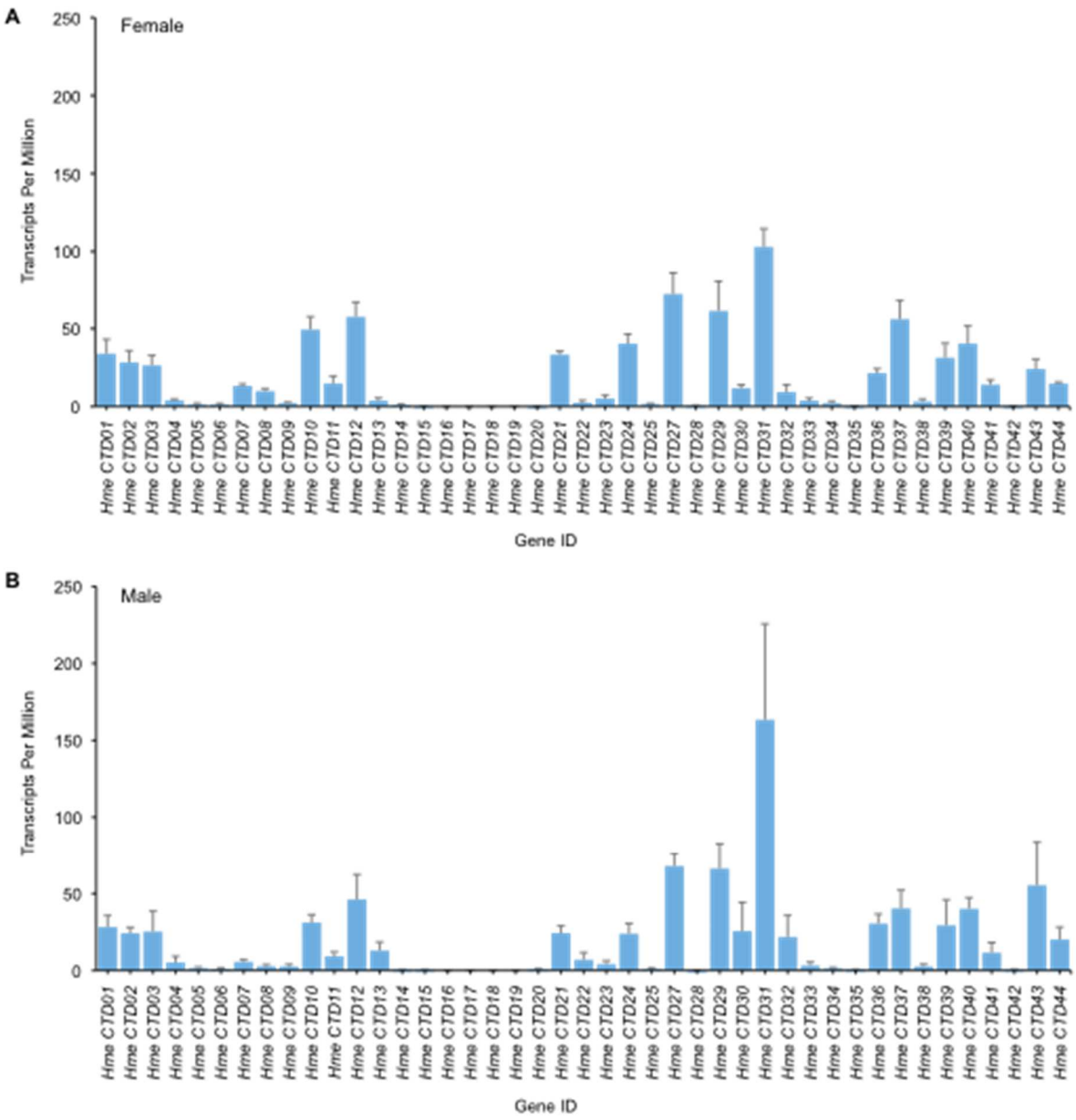


Figure S2.3. CRAL-TRIO domain gene expression in *H. melpomene* legs. (A) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* female legs with standard error bars. (B) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* male legs with standard error bars.

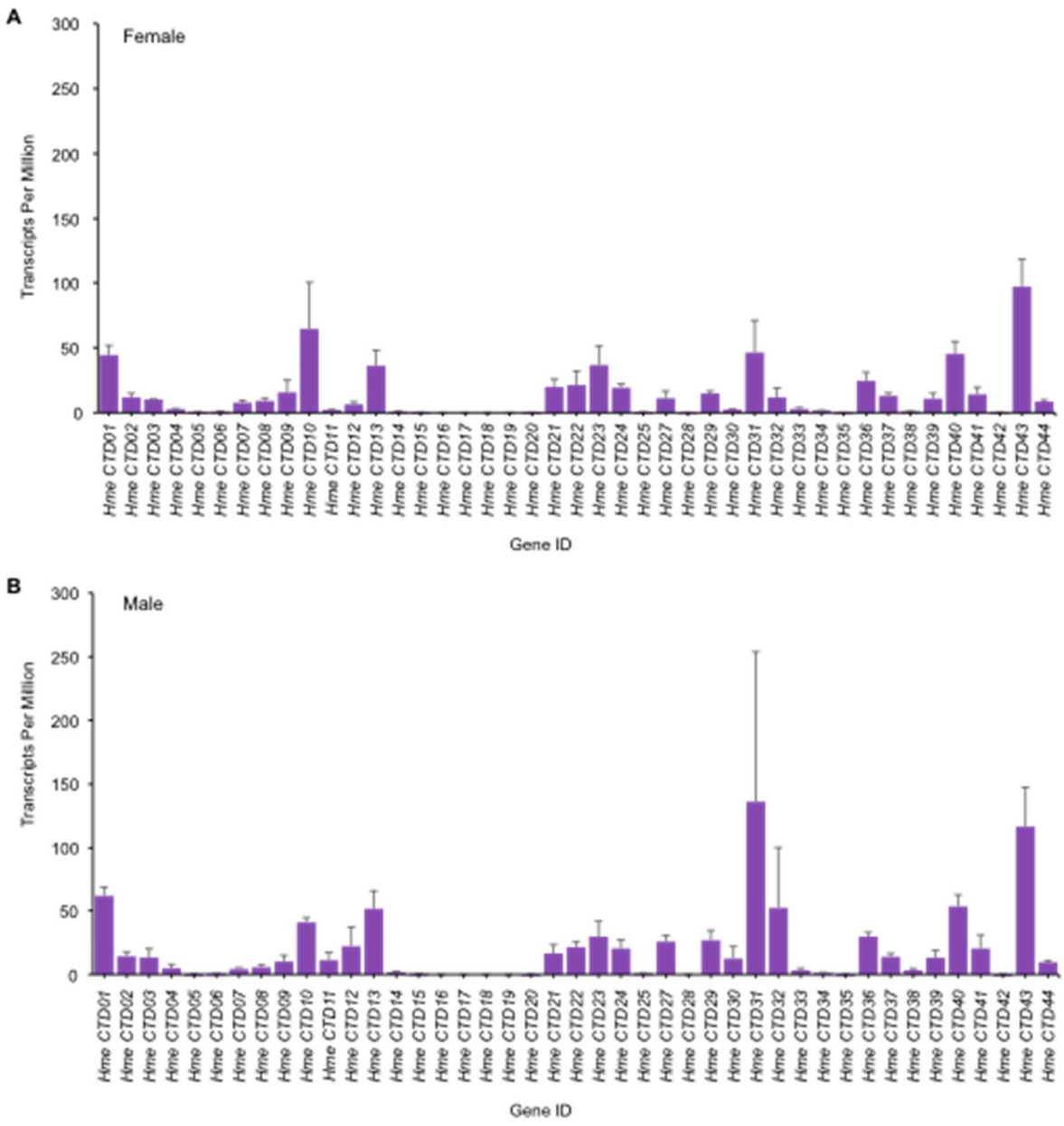


Figure S2.4. CRAL-TRIO domain gene expression in *H. melpomene* mouth. (A) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* female mouth parts with standard error bars. (B) Transcripts Per Kilobase Million (TPM) of CRAL-TRIO domain genes in *H. melpomene* male mouth parts with standard error bars.

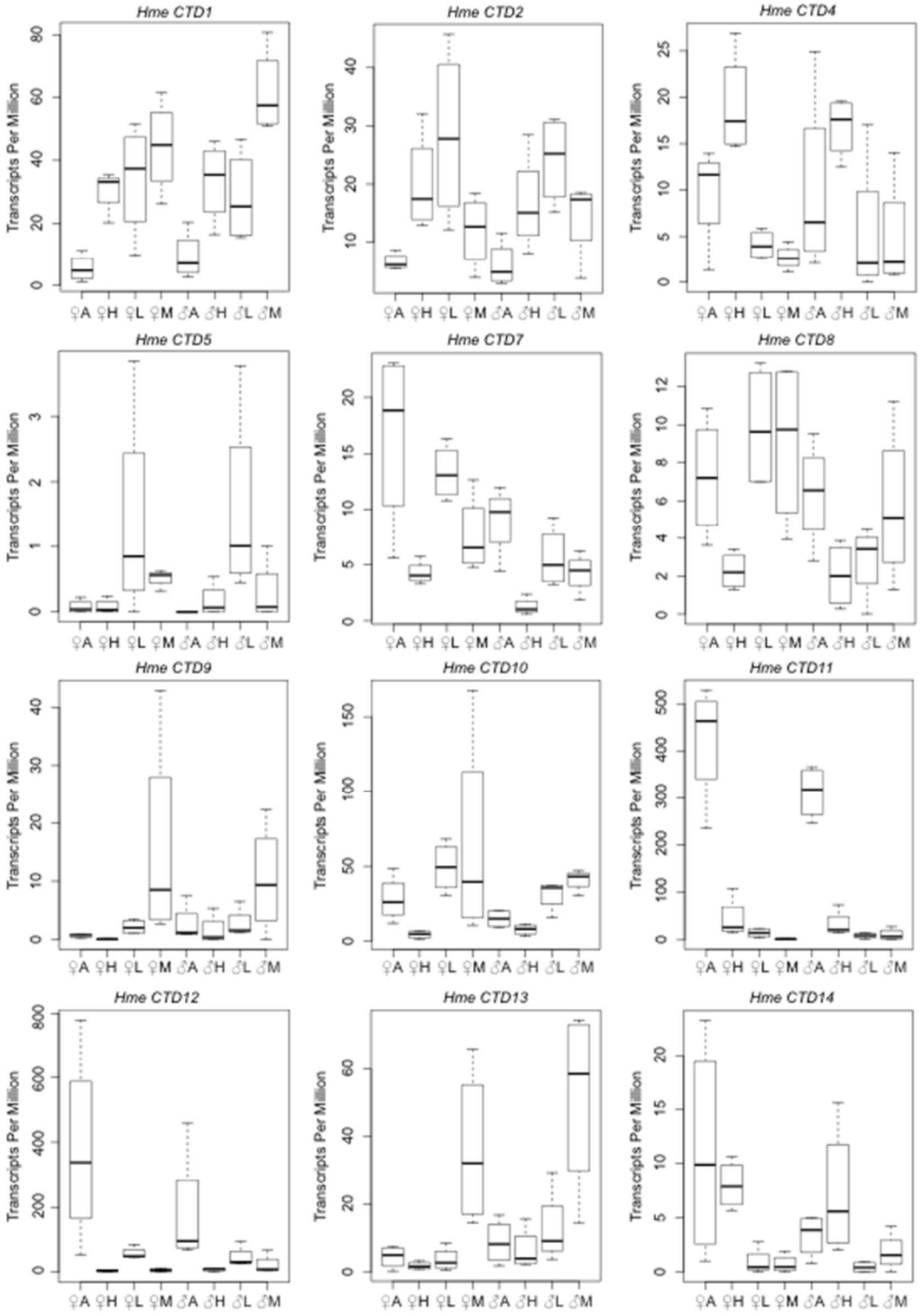


Figure S2.5. TPM plots of CRAL-TRIO domain genes (*Hme CTD1-20*) with p <0.05 using ANOVA.

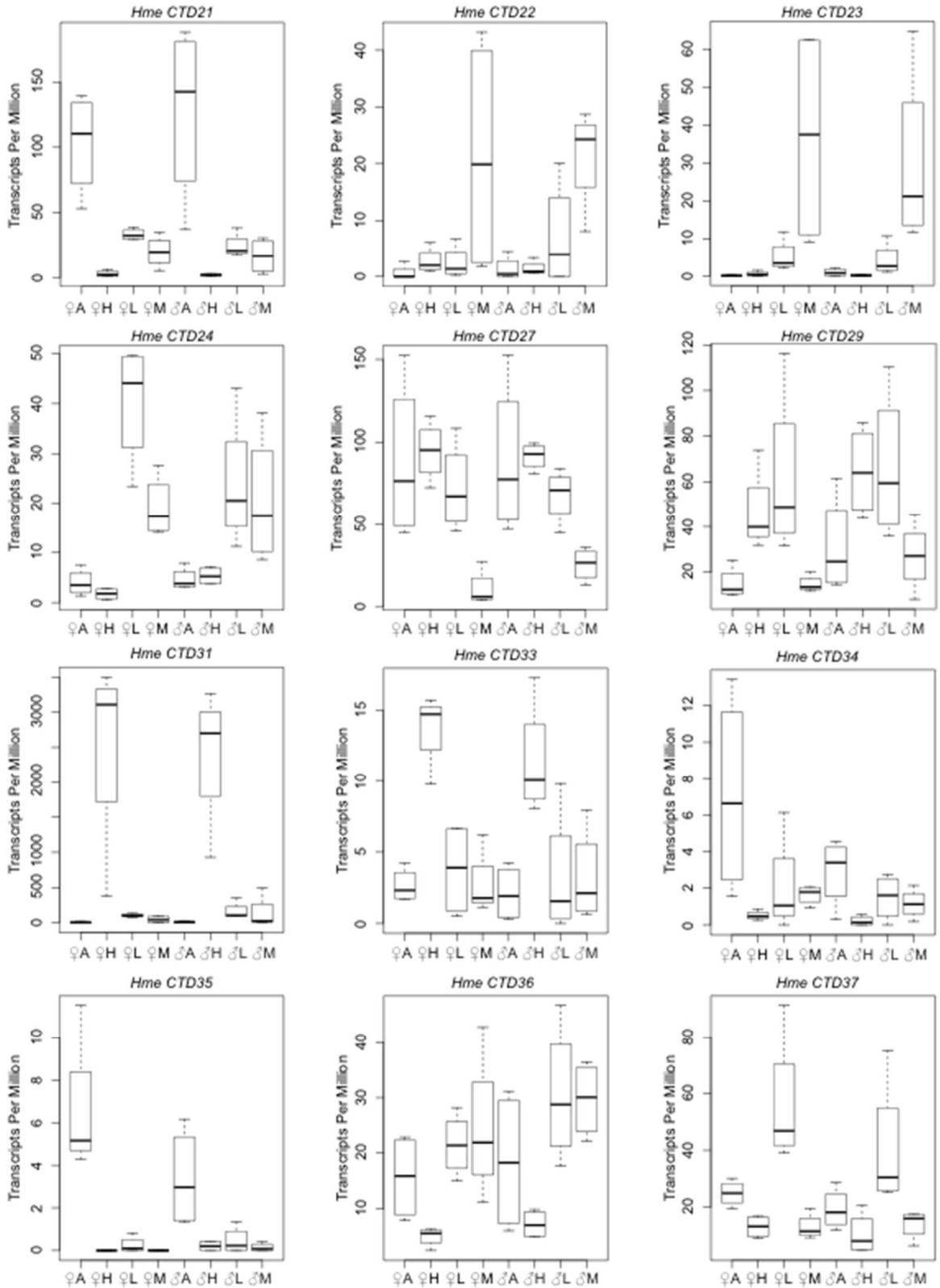


Figure S2.6. TPM plots of CRAL-TRIO domain genes (*Hme CTD21-37*) with p <0.05 using ANOVA.

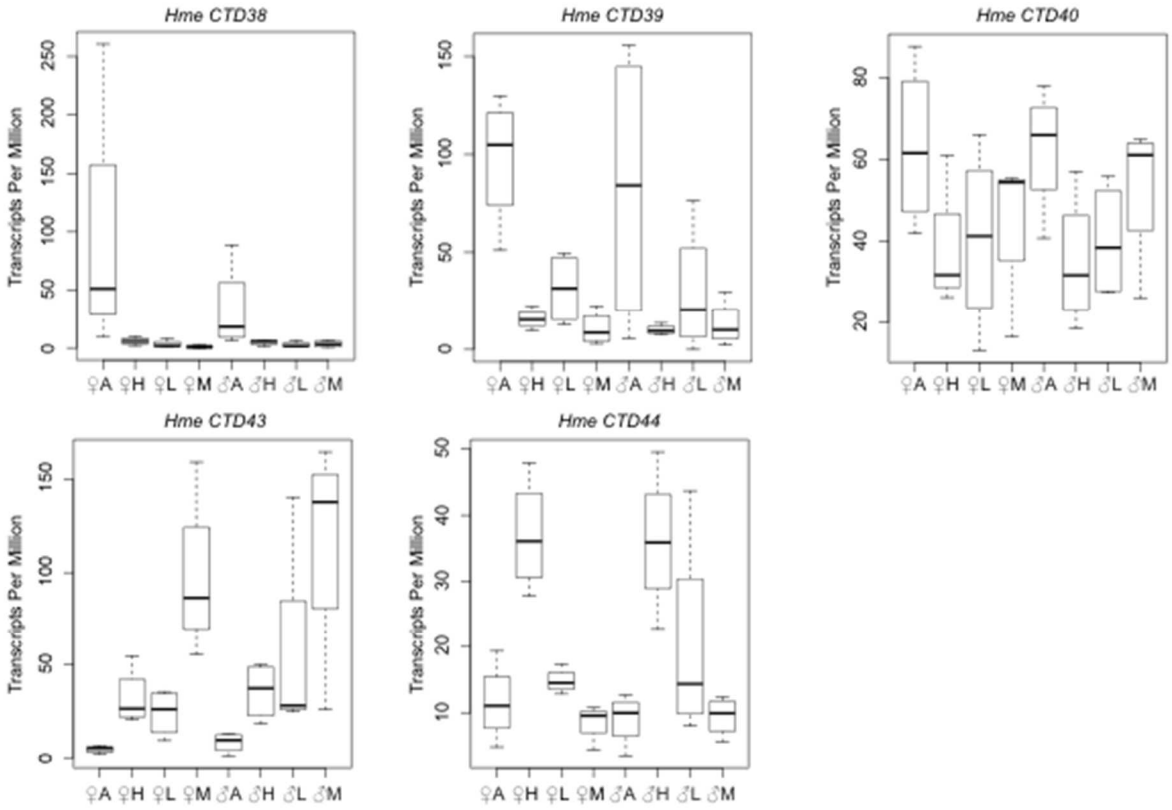


Figure S2.7. TPM plots of CRAL-TRIO domain genes (*Hme CTD38-44*) with p <0.05 using ANOVA.

Figure S2.8. Expression of CRAL-TRIO domain genes in *Bicyclus anynana*. (A) FPKM plot of CRAL-TRIO domain genes in heads of the butterfly *B. anynana*. One gene, *BA comp48725* in grey, is highly expressed and the ortholog of *Hme CTD31*. (B) Alignment of *BA comp48725* and *Hme CTD31*.

Figure S2.9. Expression of *Papilio RBP* in *H. melpomene* tissue types. FPKM plot of the *H. melpomene Papilio RBP* ortholog, *comp30064*. This gene encoding a retinol-binding protein is upregulated in *H. melpomene* male and female heads compared to other tissue types.

Table S2.1. Primer sequences.

CTD31		Start	Amplicon	Melting temperature
Forward	TCAGAGGGTGCAAGTTTTCC	164	589	60
Reverse	GGTCCAGCTTTTCCACCATA	752	589	59.9
CTD12				
Forward	TCCGAAGTCTCCAAGCAAGT	172	315	59.9
Reverse	TCCTAAAGAGACGCCGGTTA	486	315	59.8
CTD26				
Forward	TGGTTCAGCCCTGTCAATTT	73	792	60
Reverse	ATCGCTGAATTCGTTGGAAG	687	792	60

Table S2.2. CRAL-TRIO domain containing gene GenBank Accession numbers.

Gene name	Transcriptome contig ID	GenBank Accession
Hme CTD1	comp27769_c0	MG434618
Hme CTD2	comp20395_c0	MG434614
Hme CTD3	comp26992_c0	MG434616
Hme CTD4	comp29579_c0	MG434612
Hme CTD5	N/A	MG434597
Hme CTD6	N/A	MG434598
Hme CTD7	N/A	MG434599
Hme CTD8	N/A	MG434600
Hme CTD9	comp29350_c0	MG434613
Hme CTD10	comp30460_c0	MG434609
Hme CTD11	N/A	MG434601
Hme CTD12	comp27538_c0	MG434617
Hme CTD13	comp28803_c0	MG434620
Hme CTD14	N/A	MG434602
Hme CTD15	N/A	MG434603
Hme CTD16	N/A	MG434624
Hme CTD17	N/A	MG434625
Hme CTD18	N/A	MG434626
Hme CTD19	N/A	MG434627
Hme CTD20	N/A	MG434628
Hme CTD21	comp29217_c0	MG434622
Hme CTD22	comp32086_c0	MG434605
Hme CTD23	N/A	MG434629
Hme CTD24	N/A	MG434630
Hme CTD25	N/A	MG434631
Hme CTD27	N/A	MG434632
Hme CTD28	N/A	MG434633
Hme CTD29	comp29872_c0	MG434610
Hme CTD30	N/A	MG434634
Hme CTD31	comp29636_c0	MG434623
Hme CTD32	N/A	MG434635
Hme CTD33	N/A	MG434636
Hme CTD34	N/A	MG434596
Hme CTD35	comp25804_c0	MG434615
Hme CTD36	comp31868_c0	MG434606
Hme CTD37	comp28979_c0	MG434621
Hme CTD38	comp27885_c0	MG434619
Hme CTD39	comp32068_c2	MG434607

Hme CTD40	comp30622_c0	MG434608
Hme CTD41	comp32280_c0	MG434604
Hme CTD42	N/A	MG434637
Hme CTD43	comp29592_c2	MG434611
Hme CTD44	N/A	MG434638
Papilio RBP	comp30064_c0	MG434667
Ba comp28563	BA_comp28563_c0	MG434666
Ba comp36385	BA_comp36385_c0	MG434665
Ba comp40187	BA_comp40187_c0	MG434664
Ba comp41745	BA_comp41745_c0	MG434663
Ba comp42621	BA_comp42621_c0	MG434662
Ba comp42863	BA_comp42863_c0	MG434661
Ba comp43085	BA_comp43085_c0	MG434660
Ba comp43964	BA_comp43964_c0	MG434659
Ba comp44196	BA_comp44196_c0	MG434658
Ba comp46235	BA_comp46235_c0	MG434656
Ba comp46282	BA_comp46282_c0	MG434655
Ba comp46738	BA_comp46738_c0	MG434654
Ba comp46903	BA_comp46903_c0	MG434653
Ba comp47163	BA_comp47163_c0	MG434652
Ba comp47510	BA_comp47510_c1	MG434651
Ba comp47949	BA_comp47949_c0	MG434650
Ba comp48211	BA_comp48211_c0	MG434649
Ba comp48475	BA_comp48475_c0	MG434648
Ba comp48524	BA_comp48524_c0	MG434647
Ba comp48726	BA_comp48726_c1	MG434646
Ba comp49069	BA_comp49069_c0	MG434645
Ba comp49764	BA_comp49764_c0	MG434644
Ba comp50212	BA_comp50212_c0	MG434643
Ba comp50877	BA_comp50877_c0	MG434642
Ba comp51136	BA_comp51136_c3	MG434641
Ba comp52253	BA_comp52253_c0	MG434640
Ba comp9682	BA_comp9682_c0	MG434639
Ba comp44923	BA_comp44923_c0	MG434657

Table S2.3. CRAL-TRIO domain containing gene duplications.

CTD ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
CTD1	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD2	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y
CTD3	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y
CTD4	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD5	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N
CTD6	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N
CTD7	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N
CTD8	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N	N	N
CTD9	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD10	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y
CTD11	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD12	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD13	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N
CTD14	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD15	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	N	N	N	Y
CTD16	N	N	N	N	N	Y	N	N	Y	N	Y	N	Y	N	N	N	N	Y
CTD17	N	Y	N	N	N	Y	N	Y	Y	N	Y	N	Y	N	N	N	N	Y
CTD18	N	Y	N	N	N	Y	N	Y	Y	N	Y	N	Y	N	N	N	N	Y
CTD19	N	Y	N	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N	Y
CTD20	N	Y	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y
CTD21	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD22	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y
CTD23	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD24	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	Y	N
CTD25	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	Y	Y	N
CTD27	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD28	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD29	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD30	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD31	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD32	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD33	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N	N
CTD34	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD35	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD36	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD37	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD38	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD39	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

CTD40	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD41	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y
CTD42	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
CTD43	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	N	N	Y	N
CTD44	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

N is no duplication.
Y is a tandem duplication.
Rows are CRAL-TRIO domain gene IDs.
Columns are 18 resequenced genomes.

1. H. m. rosina
2. H. m. rosina
3. H. m. rosina
4. H. m. rosina
5. H. m. melpomene
6. H. m. melpomene
7. H. m. melpomene
8. H. m. melpomene
9. H. m. melpomene
10. H. m. melpomene
11. H. m. amaryllis
12. H. m. amaryllis
13. H. m. amaryllis
14. H. m. amaryllis
15. H. m. aglaope
16. H. m. aglaope
17. H. m. aglaope
18. H. m. aglaope

Table S2.4. CRAL-TRIO domain containing gene deletions.

CTD ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
CTD1	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD2	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N
CTD3	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	Y	N	N
CTD4	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD5	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD6	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD7	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
CTD8	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
CTD9	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD10	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD11	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD12	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD13	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD14	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD15	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
CTD16	N	Y	N	Y	N	N	Y	Y	Y	N	N	N	N	N	N	N	N	N
CTD17	N	Y	N	Y	N	N	Y	Y	Y	N	N	N	N	N	N	N	N	N
CTD18	N	Y	N	Y	N	N	N	N	Y	N	Y	N	N	N	N	N	N	N
CTD19	N	Y	N	N	N	N	N	N	Y	N	Y	N	Y	N	N	N	N	N
CTD20	N	Y	N	N	N	N	N	N	Y	N	Y	N	Y	N	N	N	N	N
CTD21	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
CTD22	N	N	N	N	N	Y	N	Y	N	N	N	N	N	N	Y	N	N	N
CTD23	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N
CTD24	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD25	N	N	N	Y	N	N	N	N	Y	N	Y	N	Y	Y	N	N	N	N
CTD27	N	N	N	N	N	N	N	N	Y	N	Y	N	N	N	Y	N	N	N
CTD28	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	Y	N	N	N
CTD29	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD30	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD31	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD32	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y	N	N	N	N
CTD33	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD34	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD35	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD36	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N
CTD37	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD38	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
CTD39	N	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N

CTD40	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD41	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
CTD42	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	N	Y	Y	N
CTD43	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	Y
CTD44	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

N is no duplication.
Y is a tandem duplication.
Rows are CRAL-TRIO domain gene IDs.
Columns are 18 resequenced genomes.

1. H. m. rosina
2. H. m. rosina
3. H. m. rosina
4. H. m. rosina
5. H. m. melpomene
6. H. m. melpomene
7. H. m. melpomene
8. H. m. melpomene
9. H. m. melpomene
10. H. m. melpomene
11. H. m. amaryllis
12. H. m. amaryllis
13. H. m. amaryllis
14. H. m. amaryllis
15. H. m. aglaope
16. H. m. aglaope
17. H. m. aglaope
18. H. m. aglaope

Table S2.5. Read mapping statistics.

Specimen ID	Sex	Tissue	Reads Processed	Reads Mapped	Percent Mapped	Accession
HMP110	M	head	6350702	5201583	81.91	E-MTAB-6342
HMP112	F	head	6571100	5477897	83.36	E-MTAB-6342
HMP114	M	head	6452237	5520889	85.57	E-MTAB-6342
HMP115	F	head	6433685	5497316	85.45	E-MTAB-6342
HMP333	M	head	6785156	5270272	77.67	E-MTAB-6249
HMP347	F	ant	9936121	7378930	74.26	E-MTAB-1500
HMP347	F	head	17986954	13915658	77.37	E-MTAB-6249
HMP347	F	legs	7093588	5551679	78.26	E-MTAB-1500
HMP347	F	mou	9840804	7373591	74.93	E-MTAB-1500
HMP406	M	ant	9838323	7742649	78.7	E-MTAB-1500
HMP406	M	legs	8774889	6541225	74.54	E-MTAB-1500
HMP406	M	mou	10278815	7831619	76.19	E-MTAB-1500
HMP500	M	ant	23056107	19262316	83.55	E-MTAB-1500
HMP500	M	legs	5538356	4637405	83.73	E-MTAB-1500
HMP500	M	mou	5684069	4641385	81.66	E-MTAB-1500
HMP501	M	ant	22203980	19728767	88.85	E-MTAB-1500
HMP501	M	legs	4194371	3530819	84.18	E-MTAB-1500
HMP501	M	mou	5951889	5012821	84.22	E-MTAB-1500
HMP502	F	ant	5850274	4634307	79.22	E-MTAB-1500
HMP502	F	legs	20215665	16901298	83.6	E-MTAB-1500
HMP502	F	mou	5524477	4300428	77.84	E-MTAB-1500
HMP503	F	ant	18979785	16085503	84.75	E-MTAB-1500
HMP503	F	legs	18094094	14911221	82.41	E-MTAB-1500
HMP503	F	mou	21294005	16701435	78.43	E-MTAB-1500
HMP514	F	ant	20330905	14072695	69.22	E-MTAB-6249
HMP514	F	head	22146465	16947222	76.52	E-MTAB-6249
HMP514	F	legs	21603932	16139096	74.7	E-MTAB-6249
HMP514	F	mou	22691748	15911675	70.12	E-MTAB-6249
HMP515	M	ant	18502101	14094799	76.18	E-MTAB-6249
HMP515	M	head	17019093	13289960	78.09	E-MTAB-6249
HMP515	M	legs	19851128	15479195	77.98	E-MTAB-6249
HMP515	M	mou	17081488	12353492	72.32	E-MTAB-6249

Table S2.6. All DE contigs in head vs antennae comparison.

https://academic.oup.com/gbe/article/9/12/3398/4609366#supplementary-data

Table S2.7. All DE contigs in head vs legs comparison.

https://academic.oup.com/gbe/article/9/12/3398/4609366#supplementary-data

Table S2.8. All DE contigs in head vs mouth parts comparison.

https://academic.oup.com/gbe/article/9/12/3398/4609366#supplementary-data

Table S2.9. Analysis of variance of CRAL-TRIO domain gene expression between tissue types and sexes.

	Sex		Tissue		Sex*Tissue	
Gene ID	**F Value**	**PValue**	**F Value**	**PValue**	**F Value**	**PValue**
Hme CTD1	1.059	0.314	17.686	**2.83E-06**	1.162	0.345
Hme CTD2	0.2394	0.629079	8.4428	**0.000526**	0.2613	0.85257
Hme CTD3	0.3468	0.56146	2.4334	0.08961	0.5131	0.67715
Hme CTD4	0.0291	0.865954	9.521	**0.000252**	0.2046	0.89218
Hme CTD5	0.0012	0.972794	5.095	**0.007183**	0.0949	0.962134
Hme CTD6	1.0702	0.3112	0.3083	0.8191	0.8893	0.4608
Hme CTD7	17.2952	**0.000353**	11.1007	**9.19E-05**	0.9748	0.420945
Hme CTD8	7.2135	**0.01292**	4.9067	**0.008466**	2.0718	0.130608
Hme CTD9	0.0156	0.90173	4.4716	**0.01247**	0.4077	0.74884
Hme CTD10	1.7981	0.1925	4.5493	**0.01163**	0.3649	0.77894
Hme CTD11	2.4598	0.1299	85.2287	**6.16E-13**	2.1016	0.1266
Hme CTD12	1.1153	0.301448	8.1908	**0.000629**	1.2578	0.311116
Hme CTD13	2.8245	0.1058	14.2553	**1.53E-05**	0.2697	0.8466
Hme CTD14	1.5502	0.225128	5.6029	**0.004659**	1.5169	0.235585
Hme CTD15	2.3455	0.1387	1.0876	0.3733	0.5539	0.6505
Hme CTD16	0	0	0	0	0	0
Hme CTD17	0	0	0	0	0	0
Hme CTD18	0	0	0	0	0	0
Hme CTD19	0	0	0	0	0	0
Hme CTD20	1.2101	0.2822	0.8971	0.457	0.8394	0.4857
Hme CTD21	0.0726	0.7899	24.3925	**1.82E-07**	0.5133	0.677
Hme CTD22	0.1002	0.754325	8.559	**0.000485**	0.1502	0.928538
Hme CTD23	0.1492	0.702663	10.5765	**0.000127**	0.1281	0.942475
Hme CTD24	0.8355	0.36977	21.0141	**6.72E-07**	2.4156	0.09127
Hme CTD25	0.1139	0.7386	1.3812	0.2725	0.2407	0.867
Hme CTD27	0.0441	0.8354	11.8571	**5.84E-05**	0.1949	0.8988
Hme CTD28	0.0661	0.7993	0.8984	0.4564	1.5138	0.2364
Hme CTD29	2.6075	0.119433	7.6407	**0.000938**	0.1348	0.938348
Hme CTD30	2.246	0.147	0.5294	0.6664	0.1573	0.924
Hme CTD31	0.0009	0.9757	28.5965	**4.29E-08**	0.0457	0.9867
Hme CTD32	3.4133	0.07704	2.1091	0.12558	0.6538	0.58842
Hme CTD33	0.4493	0.5091	19.2721	**1.40E-06**	0.2901	0.8321
Hme CTD34	2.7365	0.111099	5.7085	**0.004266**	1.2315	0.320008
Hme CTD35	1.5978	0.21836	20.8141	**7.29E-07**	2.5759	0.07741
Hme CTD36	2.2991	0.142508	9.4261	**0.000268**	0.2339	0.871814
Hme CTD37	1.6097	0.2167	13.7315	**2.03E-05**	0.6225	0.6074
Hme CTD38	0.9794	0.33223	3.9528	**0.02009**	1.0216	0.40053
Hme CTD39	0.2052	0.654587	10.5217	**0.000132**	0.1089	0.954066

Hme CTD40	0.0289	0.86641	3.4743	**0.03167**	0.147	0.9306
Hme CTD41	0.0421	0.8391	1.4092	0.2644	0.4223	0.7387
Hme CTD42	0.4352	0.5157	2.1963	0.1146	0.8285	0.4912
Hme CTD43	1.3993	0.2484	12.0034	**5.36E-05**	0.2975	0.8268
Hme CTD44	0.0563	0.8144	19.5543	**1.24E-06**	0.3478	0.7911

CHAPTER 3

Phototransduction in Lepidoptera

ABSTRACT

Vision is carried out by phototransduction, a signaling cascade that converts light into an electrical signal. The phototransduction cascade has not been well characterized in insect species with the exception of *Drosophila*. The order Lepidoptera provides an interesting group in which to investigate phototransduction because it encompasses diurnal butterflies and nocturnal moths. Variation in lepidopteran light environments and eye morphology might be driving divergence in the phototransduction cascade of moths and butterflies. Here we used transcriptomics and phylogenetics to identify phototransduction genes: 1) conserved between *Drosophila* and Lepidoptera, 2) specific to Lepidoptera, and 3) specific to moths or butterflies. First, we assigned functions to genes upregulated in *Heliconius melpomene* heads. We then generated 32 phylogenies encompassing 64 phototransduction genes in eight insect species to survey gene gains and losses. We found that a majority of the photototransduction genes were present in Lepidoptera and *Drosophila* as a single copy. However, we discovered an unclassified opsin, a gene loss in Lepidoptera, a Lepidoptera-specific expansion and a *Heliconius*-specific duplication. Lastly, we examined individual gene expression in heads of the moth *Manduca sexta* and in heads, antennae, legs, and mouth parts of the butterfly *H. melpomene*. Our results suggest that the *unclassified opsin* has a role in the visual system, *trpl* might have a larger role than *trp* in butterfly vision, loss of *inaE* might be compensated for by *DAGLβ-like*,

and lepidopteran paralogs *wunen1* and *wunen-like3* might have a role in lepidopteran phototransduction.

INTRODUCTION

Phototransduction is the process that underlies vision by which light information is converted into an electrical signal. Vision has intrigued scientists for centuries making phototransduction one of the best studied signaling pathways (Shichida & Matsuyama 2009). Phototransduction takes place in specialized cells known as photoreceptor cells whose membrane incorporates light-sensitive opsin proteins bound to a retinal-derived molecule (Fain et al. 2010). The most basic animal eyes are defined as a photoreceptor cell near a pigment cell (Arendt 2003). Photoreceptors are classified as ciliary or rhabdomeric depending on whether their photoreceptive membrane is derived from cilia or microvilli, respectively (Fain et al. 2010). Ciliary photoreceptors are commonly associated with vertebrate vision and rhabdomeric with invertebrate vision. Yet, exceptions exist such as jellyfish, which are primitive invertebrates and have ciliary photoreceptors (Suga et al. 2008). While a marine ragworm *Platynereis dumerilii* has rhabdomeric photoreceptors in its larval and adult eyes, it has ciliary photoreceptors in its brain that expressed an opsin whose sequence clusters with vertebrate opsins (Arendt 2004). Furthermore, ciliary opsins have been identified in 14 out of 27 insect genomes surveyed (Feuda et al. 2016). Differences in photoreceptor structure are accompanied by differences in the phototransduction cascade that proceeds when light is detected by ciliary and rhabdomeric photoreceptors (Fain et al. 2010). Here, we focus on *Drosophila* phototransduction which serves as a model for insect vision and the rhabdomeric phototransduction cascade.

The genes and proteins involved in *Drosophila* phototransduction have been experimentally tested for over 40 years (Hardie 2001; Hardie & Raghu 2001; Katz & Minke 2009; Montell 2012; Hardie & Juusola 2015). Phototransduction begins when light is absorbed by a chromophore molecule (11-*cis*-3-hydroxyretinal in butterflies) bound to a 7-transmembrane opsin protein forming a light sensitive pigment called rhodopsin. A single photon of light causes the chromophore to change its confirmation from *cis* to all-*trans* (von Lintig et al. 2010). This change in configuration triggers a G-protein-coupled cascade (Figure 3.1) that activates a phospholipase C (PLC) (Bloomquist et al. 1988). PLC hydrolyses phosphatidylinositol 4,5-bisphosphate (PIP_2) to produce inositol 1,4,5-trisphosphate ($InsP_3$) and diacylglycerol (DAG) (Bloomquist et al. 1988; Hardie 2001). Concurrently, by a mechanism that is not well understood, there is an opening of Ca^{2+}-permeable light-sensitive transient receptor potential channels (TRP) and TRPL (transient receptor potential-like) which causes depolarization of the cell (Montell & Rubin 1989; Hardie & Minke 1992; Niemeyer et al. 1996; Shieh & Zhu 1996; Montell 2005). Finally, phototransduction is terminated when the active rhodopsin (metarhodopsin) binds arrestin (Dolph et al. 1993; Stavenga & Hardie 2011). Although phototransduction in *Drosophila* is the most extensively investigated among insects, a study of vision-related genes in four insect genomes (mosquito, red flour beetle, honeybee and *Drosophila*) found gains and losses of genes involved in phototransduction across different linages (Bao & Friedrich 2009). *Drosophila* had by far the largest number of gene gains compared to the other insects examined. This implies that other insect species might differ significantly in the genes underlying their phototransduction.

Phototransduction cascades have not been described in other insects. Numerous studies have focused on characterizing the opsins expressed in photoreceptor cells and their arrangement across the compound eye (Spaethe & Briscoe 2005; Henze et al. 2012; Futahashi et al. 2015; McCulloch et al. 2016; Perry et al. 2016; Giraldo-Calderón et al. 2017; McCulloch et al. 2017). Although a large focus is on the opsins, changes in the downstream pathway by which opsins function might also contribute to differences in visual systems (Plachetzki et al. 2010). Fewer studies have probed other aspects of the cascade, using phylogenetic trees or gene expression to decipher whether phototransduction genes have conserved functions in different insect species. Messenger RNA (mRNA) extracted from the troglobiont beetle, *Ptomaphagus hirtus*, recovered 20 phototransduction genes expressed in adult heads (Friedrich et al. 2011). Similarly, expression of opsins and TRP channels were quantified in the nocturnal cockroach to determine whether insects in different light environments vary in their phototransduction mechanism (French et al. 2015). mRNA showed that one green opsin and TRPL were more highly expressed in *P. americana* relative to other opsins and TRP, and RNAi (RNA interference) of these genes resulted in electroretinogram (ERG) reduction (French et al. 2015). A recent study exposed the oriental armyworm, *Mythimna separate* (Noctuidae), to different light environments and found differential expression of phototransduction genes in adult heads (Duan et al. 2017). Phototransduction genes were also differentially expressed between seasonal forms in heads of the butterfly *Bicyclus anynana* (Macias-Muñoz et al. 2016). Lepidoptera, moths and butterflies, provide a unique group with which to investigate the molecular evolution and expression of phototransduction genes in insects adapted to different light

environments (Yagi & Koyama 1963; Horridge et al. 1972; Nilsson et al. 1984; Yack et al. 2007; Warrant & Dacke 2016).

Phylogenetic trees and gene expression analyses of phototransduction genes in Lepidoptera may reveal: 1) the extent to which aspects of *Drosophila* phototransduction are conserved in Lepidoptera, 2) possible lepidopteran-specific phototransduction features, 3) variations between diurnal and nocturnal Lepidoptera. Phylogenies uncover gene loss events and the emergence of recent gene duplications. In insects, phylogenetic analyses have been used to reveal duplications of opsin genes that had not been previously described (Spaethe & Briscoe 2004; Briscoe et al. 2010). A survey of 23 vision-related gene families in 19 metazoan genomes revealed that eye development and phototransduction genes have higher rates of retention and duplications in pancrustaceans (Rivera et al. 2010). Since only the nocturnal domesticated silkmoth *Bombyx mori* was used in the pancrustacean study and only five gene families involved in phototransduction were examined (r-opsin, TRP, phospholipase C, Gq-alpha and arrestin) (Rivera et al. 2010), it remains to be seen if there are additional differences in phototransduction genes between *Drosophila* and moth and butterfly species. In our present study we expand on the genes surveyed thus far by looking at 64 phototransduction-related genes.

While gene phylogenies tell us the probable molecular history of gene families, gene expression data is a first step towards inferring gene function. Genes involved in vision should be highly expressed in photoreceptor cells and upregulated in the eyes relative to other tissue types, thus visualizing or quantifying where phototransduction genes are expressed will reveal whether they have a potential role in vision. As an example, the horseshoe crab *Limulus polyphemus* has 18 opsins, some of which are expressed only in the

eyes, in eyes and central nervous system, exclusively in the central nervous, and some are not expressed in either (Battelle et al. 2016). It is possible that the opsins missing from the eyes and central nervous system are expressed in other tissue types and have other non-visual functions (Feuda et al. 2016) or are not expressed at all. In *H. melpomene*, the reference genome (Davey et al. 2016) reveals a *UVRh* duplication but mRNA shows that one of the copies is downregulated in this species and only one of the copies has protein expression in the compound eye (McCulloch et al. 2017). These studies highlight the importance of measuring gene expression in candidate tissues before inferring gene function based on sequence. Further, it is also possible that a non-ortholog member of the same gene family can be co-opted to partake in the predicted visual function. As an example, an expression analysis of CRAL-TRIO domain genes in *H. melpomene* found that an ortholog of *Drosophila pinta* is missing in Lepidoptera and instead a paralog carries out a similar function of chromophore binding (Wang & Montell 2005; Smith & Briscoe 2015; Macias-Muñoz et al. 2017). Moreover, as observed in the cockroach, while genes such as *TRP* and *TRPL* are conserved and expressed, one gene copy might have a greater impact on phototransduction than the other (French et al. 2015). Consequently investigating both gene gain/loss and the expression of phototransduction genes in Lepidoptera might uncover differences in their visual processing that helps them specialize to different light environments.

In this study we combined transcriptomics and phylogenetics to perform the first investigation of phototransduction genes in Lepidoptera. We used RNA-Sequencing data from four tissue types of the butterfly *Heliconius melpomene* to identify genes upregulated in heads. We hypothesized that genes upregulated in heads had eye and vision-related

functions so we annotated these genes with *Drosophila* gene IDs to uncover phototransduction genes with conserved function. We also extracted phototransduction gene sequences from reference genomes for *Drosophila melanogaster*, *Anopheles gambiae*, *Apis mellifera*, *Tribolium casteum*, *Bombyx mori*, *Manduca sexta*, *Danaus plexippus* and *Heliconius melpomene*. In case any genes were missing annotations in the reference assemblies, we searched *de novo* transcriptome assemblies for *M. sexta, H. melpomene* and *D. plexippus*. We generated 32 phylogenetic trees for 64 phototransduction genes to identify gene gain or loss between *Drosophila* and Lepidoptera and within Lepidoptera, between moths and butterflies. We found that most genes were conserved between *Drosophila* and Lepidoptera with some exceptions (Figure 3.1). For gene families with gene loss or gain, we quantified the expression of paralogs in *M. sexta* heads and in four tissue types of *H. melpomene* to detect genes expressed in heads that might have a role in vision. Our findings suggest that similar genes partake in the phototransduction cascade of moths and butterflies, and that there is variation between Lepidoptera and *Drosophila* in instances where Lepidopteran-specific paralogs take on a function in vision.

MATERIALS AND METHODS

Transcriptome-wide differential expression analysis

RNA-Sequencing data for *H. melpomene* male and female head, antennae, legs and mouth parts were obtained from array express projects E-MTAB-1500 and E-MTAB-6249 (Table S3.1). A four tissue *de novo* transcriptome made from one library per tissue type per sex was used as reference (doi:10.5061/dryad.857n9; see Macias-Muñoz et al. 2017). Reads from each sample were mapped to the transcriptome using bwa (Li & Durbin 2009)

and RSEM (Li & Dewey 2011) was used to quantify mapped raw reads. We used edgeR (Robinson et al. 2010) to perform three pairwise comparisons for differential expression analysis: head versus antennae, head versus legs, and head versus mouth parts. For each comparison, a generalized linear model was used to include terms for batch, tissue, sex, the interaction of sex and tissue (~batch + tissue + sex + sex*tissue). Each analysis also included filtering to remove lowly expressed contigs (less than 1 count per million for at least 4 groups). Samples were normalized using a trimmed mean of the log expression ratios (TMM) (Robinson & Oshlack 2010). After each comparison, p-values were further corrected using a Bonferroni false discovery rate (FDR) correction. Contigs were considered significantly differentially expressed when the FDR was less than 0.05 and the log fold change (logFC) was greater than 1.

Of these differentially expressed contigs, we identified which were upregulated in heads for each comparison. The resulting gene lists were merged to identify contigs commonly upregulated in heads. Patterns of expression for significant contigs and those commonly upregulated in heads were visualized using heatmaps (Ploner 2012). Contigs were annotated with *Drosophila* gene IDs (Marygold et al. 2013) by using command-line BLAST+ to compare *H. melpomene* transcriptome sequences to *Drosophila* gene sequences (Camacho et al. 2009). We used batch download in Flybase to acquire gene ontology terms (GO terms) for our differentially expressed and head upregulated contigs. Differentially expressed contigs with unique annotations were enriched for function using a Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al. 2009). Contigs commonly upregulated in heads were also assigned GO terms and protein classification NCBI Blast and InterProScan in BLAST2GO to uncover additional annotations potentially

missing from a comparison to *Drosophila* only (Conesa et al. 2005; Conesa & Götz 2008; Götz et al. 2008).

Phototransduction genes in insect genomes

To identify phototransduction genes in Lepidoptera and explore their evolutionary history, we used experimentally tested *D. melanogaster* sequences to search for homologs in published insect genomes. We began with a compilation of sequences by Bao and Friedrich (Bao & Friedrich 2009) but expanded it to include Lepidoptera and additional phototransduction genes (Table S3.2). We used BLAST to search the genomes of *Anopheles gambiae*, *Apis mellifera*, *Tribolium casteum*, *Bombyx mori, Manduca sexta*, *Heliconius melpomene,* and *Danaus plexippus*. Sequences with identity of more than 20% and an e-value greater than 1E-10 were tested for homology using reciprocal blastp to the NCBI database. In addition to searching Lepidoptera reference genomes, we searched *de novo* transcriptome assemblies to improve annotation and find duplicates that are not found in genomes. We searched an *H. melpomene* four tissue transcriptome (doi:10.5061/dryad.857n9; Macias-Muñoz et al. 2017) and a *M. sexta* head transcriptome (doi:10.5061/dryad.gb135; Smith et al. 2014). The nucleotide sequences recovered from *de novo* transcriptomes were translated using OrfPredictor with the blastx option before testing them by reciprocal blast hits (Min et al. 2005).

Sequence corrections were accomplished by aligning sequences in Molecular Evolutionary Genetics Analysis (MEGA) and manually correcting missing pieces then using BLAST to recover the segment from the genome. To obtain the consensus sequences, we inputted corrected sequences to CLC Genomics (CLCBio) and mapped reads against them.

With some exceptions, we recovered the entire sequence for all phototransduction genes in *H. melpomene* and *M. sexta*. Phototransduction genes for *H. melpomene* and *M. sexta* were annotated and deposited in GenBank with Accession numbers: XXXXXXXX-XXXXXXXX. We also searched NCBI for the insect sequence matches to our annotated transcriptome sequences, and added any hits missing from our data set. In addition, to test the history of *inaE* gene in *Drosophila* and the lepidopteran *DAGLβ-like* in the context of the evolution of the gene family in animals, we added *Homo sapiens* and *Mus musculus* homolog sequences using blastp searches of NCBI data bases for *H. sapiens* (taxid:9606) and *M. musculus* (taxid:10090).

Protein sequences for each gene family were aligned in MEGA 7.0 using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004; Kumar et al. 2016). The alignments were further corrected manually. Before generating maximum likelihood trees, we calculated Bayesian Information Criterion (BIC) values to assess which substitution model would best fit our data (Schwarz 1978; Kumar et al. 2016). We used the best fit model to generate phylogenies using 100 bootstrap replicates.

Expression of candidate genes

In order to study expression patterns among homologs, we looked at the expression of genes belonging to five gene families in *M. sexta* heads and in *H. melpomene* heads, antennae, legs, and mouth parts (labial palps + proboscis). We began by adding our corrected *H. melpomene* and *M. sexta* sequences to the *de novo* transcriptome assembly. We mapped trimmed and parsed reads from four male and four female *M. sexta* heads (E-MTAB-2066; Smith et al. 2014) to the corrected *M. sexta* transcriptome. We also mapped

processed reads from *H. melpomene* head, antennae, legs and mouth parts (E-MTAB-1500, E-MTAB-6249, E-MTAB-6342; Macias-Muñoz et al. 2017) to the corrected *H. melpomene* transcriptome. RSEM was used to count raw reads mapped (Li & Dewey 2011). We visualized expression levels by graphing Transcripts Per Million (TPM) for each gene of interest using ggplot2 (Wickham 2009).

RESULTS AND DISCUSSION

Transcriptome-wide differential expression analysis

In order to determine the functions of genes expressed in butterfly heads, we used *H. melpomene* RNA-Seq data to identify contigs upregulated in head tissue relative to antennae, legs, and mouth parts. A multidimensional scaling (MDS) plot showed that head libraries group together and away from other tissue types (Figure 3.2A). Differential expression analysis comparing heads vs. antennae yielded 1,173 Differentially Expressed (DE) contigs (Figure S3.1; Table S3.3), 561 of these were upregulated in heads (Table 3.1). Analysis of head vs. legs mRNA gave 1,472 DE contigs (Figure S3.1; Table S3.4), of these contigs 928 were upregulated in heads. Heads vs. mouth parts comparison yielded 1,486 DE contigs (Figure S3.1; Table S3.5), 914 of these were upregulated in heads (Table 3.1). DE contigs from each of the three pairwise comparisons matched 590, 748, and 700 unique gene ontology terms (GO terms; Table 3.1).

We performed an enrichment analysis for contigs DE between the 3 comparisons (head vs. legs, head vs. antennae, and head vs. mouth parts) to investigate the potential functions of these genes. We found that DE contigs for the three comparisons had some similar functions. Annotation terms that were similar across the three comparisons

included detection of light stimulus, regulation of rhodopsin mediated signaling, and homeobox domain (Table S3.6). The first two annotation clusters included genes involved in phototransduction. The homeobox cluster included genes involved in antennal, leg and neuron development, as well as genes involved in compound eye development and morphogenesis such as *araucan*, *PvuII-PstI homology 13*, *ocelliless*, and *eyegone*. An annotation term unique to the head vs. antennae comparison was glucose-methanol-choline oxidoreductase which included the genes *glucose dehydrogenase* and *ninaG* among other yet unnamed genes (Table S3.6). An annotation term unique to the head vs. mouth parts comparison was ion channel activity and included genes involved in perception of touch, taste, and olfaction (Table S3.6). This cluster also included genes potentially involved in phototransduction such as *cacophony*, *NMDA receptor 1*, and *transient receptor potential-like* (*trpl*; Table S3.6).

Most of the genes enriched in the DE analyses between heads and other tissues are biased towards vision, as has also been found in a recent transcriptomic analysis of *M. sexta* adult head tissue alone (Smith et al. 2014). This could be because more transcription is actively occurring in the adult butterfly head and the head is mostly composed by the eye and optic lobe (Girardot et al. 2006). *Heliconius* butterflies have large eyes due to selective pressures that favor development of large eyes regardless of body size and the optic lobe accounts for approximately 64% of the total brain volume (Seymoure et al. 2015; Montgomery et al. 2016).

Head upregulated genes

We merged the lists of contigs upregulated in heads in each pairwise comparison to obtain 281 contigs commonly upregulated in heads across the three comparisons (Figure 3.2B; Table 3.2). Head upregulated contigs annotated using BLAST2GO level 2 analysis showed that 71 of the annotated genes were involved in cellular processes and 32 were involved in response to stimulus (Figure 3.2C) (Conesa et al. 2005; Conesa & Götz 2008; Götz et al. 2008, 2011). A breakdown of these genes shows that a majority are involved in ion transmembrane transport and G protein coupled receptor signaling pathway (Figure 3.2D).

Commonly upregulated contigs in heads across the three comparisons corresponded to 154 unique *Drosophila* GO terms (Table 3.1; Table S3.6). These contigs clustered into nine annotation clusters using the highest stringency in DAVID (Figure 3.2E; Huang et al. 2009). The top three annotation clusters were: 1) detection of light stimulus, 2) regulation of rhodopsin mediated signaling pathway and 3) detection of light stimulus involved in visual perception (Figure 3.2E). The genes grouped within these annotation clusters encode proteins involved in phototransduction (Figure 3.2E). Of the remaining eight annotation clusters, clusters 9 and 10 are also directly associated with vision and are enriched for homeobox and rhabdomere development, respectively. Two genes in common between these two clusters include *PvuII-PstI homology 13* (*Pph13*) and *ocelliless* (*oc*) that function in ocellus and compound eye photoreceptor development (Fichelson et al. 2012; Mahato et al. 2014).

Some of the genes enriched in other annotation clusters have a role in vision. One gene in common between annotation clusters 4, 5 and 6 is *ora transientless* (*ort*), a gene

that is necessary for vision as it encodes for a channel that receives inputs from photoreceptors (Gengs et al. 2002). Annotation clusters 4, 5 and 8 include *resistant to dieldrin* (*Rdl*), a gene that has a role in the circuits underlying visual processing, odor coding, learning and memory, sleep and courtship behavior (Brotz et al. 2001; Liu et al. 2007; Chung et al. 2009; Yuan et al. 2013).

Phototransduction genes

Genes commonly upregulated in heads were annotated with functions relating to vision and phototransduction (Figure 3.1E). Yet their evolutionary history and potential functional conservation requires further validation. To evaluate whether phototransduction genes were lost or expanded in Lepidoptera relative to *Drosophila*, we generated 32 insect phylogenies for 64 phototransduction-related genes (Table S3.2). Of the genes that we searched for, we found 45 of them in both Lepidoptera and *Drosophila*. Some of the differences that we identified between *Drosophila* and Lepidoptera were loss of *inactivation no afterpotential E* (*inaE*) and *inactivation no afterpotential C* (*inaC*) in Lepidoptera. The genes we found in Lepidoptera but not Drosophila that we identified were *pteropsin*, *unclassified opsin-like*, *ninaC2* (*neither inactivation nor afterpotential C*), *DAGL-beta*, 3 copies of *wunen* (*wun*), 3 copies of *wunen-like*, and *Vha100-1-like* (see below). There were no consistent differences between moths and butterflies, at least in the genes we surveyed. Within Lepidoptera some of the differences that we identified were a *UVRh* duplication in *H. melpomene* (below), a cerebral *LWRh* in *B. mori*, and an absence of an unclassified opsin in *B. mori*. To visualize which lepidopteran paralog had a potential

function in vision, we aligned reads to members of six gene families from *M. sexta* heads and *H. melpomene* four tissues sequencing data.

Opsins in Lepidoptera

We began our survey of phototransduction genes in Lepidoptera by investigating the molecular evolution and expression of opsin genes. Opsin phylogenies have been the focus of many studies attempting to understand the evolutionary history of light detection (Arendt 2003; Raible et al. 2006; Plachetzki et al. 2007; Suga et al. 2008; Porter et al. 2012; Ramirez et al. 2016; Vöcking et al. 2017). These studies have reconstructed opsin presence in the ancestor of bilaterian animals (Ramirez et al. 2016) and have described a new opsin type (Vöcking et al. 2017). In our transcriptome-wide analysis with *Drosophila* annotations, we found homologs of *Drosophila* rhodopsin genes *Rhodopsin 3* (*Rh3*) and *Rhodopsin 5* (*Rh5*), which correspond to *UVRh1* and *BRh* respectively, and *LWRh* upregulated in *H. melpomene* heads (Table S3.5) (Briscoe et al. 2010; Yuan et al. 2010). We also found a homolog of *neither inactivation nor afterpotential A* (*ninaA*) upregulated in *H. melpomene* heads (Table S3.5). In *Drosophila*, a mutation in *ninaA* results in a reduction of Rh1, the rhodopsin protein found in cells R1-R6 in *Drosophila* (Shieh et al. 1989). To inspect the phylogenetic history of the opsins, we added *H. melpomene* sequences from the reference genome and a *de novo* transcriptome to a set of sequences used in Kanost et al. (2016). We recovered the previously described *Heliconius*-specific *UVRh* duplication and orthologs for all other known opsins (Briscoe et al. 2010; Yuan et al. 2010; McCulloch et al. 2017). We also found *H. melpomene* and *D. plexippus* transcripts for an unclassified opsin first described in Kanost et al. (2016) for *M. sexta* (Figure 3.3A).

In order to determine whether the opsin genes recovered maintained a role in vision, we examined their expression in *M. sexta* heads and in the head, antennae, legs and mouth parts of *H. melpomene*. In *M. sexta*, all opsins present in the species were expressed in head tissue (Figure 3.3B). In *H. melpomene*, all opsins were more highly expressed in heads relative to other tissue types in accordance with the transcriptome-wide approach (Figure 3.3). *LWRh* was the most highly expressed opsin gene which corresponds to the amount of LW photoreceptor cells per ommatidium. *Heliconius* have 9 photoreceptor cells in a flower-like arrangement where six cells express *LWRh* and two express short wavelength *BRh*, *UVRh1* or *UVRh2* (McCulloch et al. 2016, 2017). In *M. sexta*, *Rh7* was the most highly expressed opsin gene (Figure 3.3B). *Rh7* plays a role in phototransduction and circadian clocks in *Drosophila* (Ni et al. 2017). The function of *Rh7* in other species has not been established but it is ancient and widespread among insects (Feuda et al. 2016). Interestingly, the unclassified opsin had expression comparable to the known visual opsins for *H. melpomene* and *M. sexta*. Kanost et al (2016) noted the unclassified opsin lacks a lysine at the typical location where the chromophore is bound in opsins, while a recent study found that alternative amino acids sites may be used in some GPCRs for chromophore-binding (Faggionato & Serb 2017). Use of expression data allowed us to determine that the unclassified opsin was upregulated in *H. melpomene* heads across the three pair-wise comparisons suggesting it may play a role in vision (Table S3.2-4; Table S3.6).

Loss of inaE

After phototransduction is triggered by photon absorption, the G-coupled protein domain activates phospholipase C (PLC encoded by *norpA*) to produces diacylglycerol

(DAG) (Bloomquist et al. 1988). G proteins mediate phototransduction when $G_q\alpha$ is released from a G-protein complex of 3 subunits (α, β, and γ) and activates phospholipase C (PLC) (Lee et al. 1994). We found homologs of *G protein α q subunit* (*Galphaq*), *G protein β-subunit 76C* (*Gbeta76C*), and *norpA* (*no receptor potential A*), upregulated in *H. melpomene* heads using a transcriptome-wide approach. DAG and its potential metabolite, polyunsaturated fatty acid (PUFA), have been implicated in the activation of transient receptor potential (TRP and TRPL) channels (Chyb et al. 1999; Leung et al. 2008). In *Drosophila*, another mechanism by which DAG is hydrolyzed is by the actions of DAG lipase (DAGL) encoded by the gene *inaE* (Leung et al. 2008). *InaE* was the only phototransduction gene that was lost in Lepidoptera (Figure 3.4A). *InaE* mutants in *Drosophila* have defective responses to light, demonstrating that DAGL activity is required for photoreceptor responses (Leung et al. 2008).

Although *inaE* was lost in Lepidoptera, we found a *DAGL β* gene in the lepidopteran genomes. To assess the relationship of *DAGL-like* to *inaE*, we included *DAGLα* and *β* sequences from insects and mammalian *H*. *sapiens* and *M. musculus*. We found that Lepidoptera and *T. castaneum* retain *DAGLβ*, *Drosophila* only retains *DAGLα* (*inaE*) and *A. mellifera*, *A. gambiae* and mammals retain both (Figure 3.4A). We predicted that *DAGLβ* carries out the phototransduction function of hydrolyzing DAG in moth and butterfly vision because Lepidoptera have lost an ortholog of *Drosophila inaE* and retained *DAGLβ*. *DALGβ* was expressed in *M. sexta* heads and in *H. melpomene* heads, antennae, legs, and mouth parts (Figure 3.4B). While we confirm expression in heads, *DAGLβ* is not upregulated in heads relative to other tissue types thus a strict function in vision for this gene is not supported.

TRP and TRPL

Transient receptor potential (TRP and TRPL) channels are essential in *Drosophila* phototransduction to allow the influx of Ca^{2+} and cause cell depolarization (Montell & Rubin 1989). *Trp* is concentrated in the rhabdomeres and flies with mutated *trp* behave as though they are blind (Montell & Rubin 1989). Since the molecular characterization of the *trp* locus, more members of the gene family have been identified. The TRP superfamily contains more than 20 cation channels (Montell et al. 2002). While *trp* and *trpl* function in *Drosophila* vision, other *trp* genes sense pain, vanilloid compounds, and heat, among other stimuli (Montell et al. 2002; Montell 2005). Another gene *trpgamma* encodes a protein that is found in Drosophila photoreceptors and forms a heteromultimeric channel with TRPL (Montell 2005).

In our examination of the TRP gene family we searched for gene sequences and calculated the expression of *trp*, *trpl*, and *trpgamma*. We found *trp*, *trpl*, and *trpgamma* as single copies in all 8 of our insect genomes (Figure 3.5A). All three genes belonging to the TRP family were expressed in *M. sexta* heads and in *H. melpomene* four tissues (Figure 3.5B). Yet, *trpl* was the most highly expressed *trp* gene in *M. sexta* heads and was the only *trp* gene upregulated in *H. melpomene* heads using a transcriptome-wide approach (Figure 3.5A; Table S3.2). A transcriptome study done in cockroaches found that *trpl* was approximately 10 times more abundant than *trp* (French et al. 2015). RNAi of *trpl* reduced electroretinogram (ERG) response much more than *trp* after 21 days suggesting that, unlike in *Drosophila*, in cockroaches TRP and TRPL do not have similar contributions to

phototransduction (French et al. 2015). Similarly, our results suggest that in lepidopteran vision *trpl* might play a larger role than *trp*.

Two additional genes that were upregulated in *H. melpomene* heads and encode proteins that interact with the TRP family are *CalX* and *inaD*. *Drosophila* has three Na+/Ca^{2+} exchangers, one of which is encoded by *CalX* (Wang et al. 2005). Mutations of *CalX* result in a transient light response and a decrease in signal amplification implying a role for this gene in Ca^{2+} maintenance for proper TRP signaling (Wang et al. 2005). Some proteins such as TRPL, Gq and Arr1 and Arr2 undergo light-dependent movement in and out of the rhabdomeres (Montell 2012). *InaD* is required to localize and coordinate proteins in the phototransduction cascade to the microvillar membrane (Bähner et al. 2000). INAD forms a complex with TRP, PLC and protein kinase C (PCK), this complex is a target of Gαq (Shieh et al. 1989; Chevesich et al. 1997; Tsunoda et al. 1997; Bähner et al. 2000; Montell 2005). The expression of *Calx* and *inaD* in *H. melpomene* suggest these genes have a role in butterfly phototransduction, potentially through similar interactions with TRP as in *Drosophila*.

Expansion of wunen in Lepidoptera

The opening of the TRP channel is also controlled by DAG level regulation by degeneration A (RDGA) and Lazaro (LAZA) (Garcia-Murillas et al. 2006; Bao & Friedrich 2009). Laza is a lipid phosphate phosphatase (LPP) and is found in *Drosophila* photoreceptors (Garcia-Murillas et al. 2006). *Lazaro* is a member of the *wunen* subfamily (Figure 3.6A). *Wunen* helps regulate the level of bioactive phospholipids, has a role in germ line migration, and is necessary for tracheal development (Zhang et al. 1997; Ile et al. 2012). We found 7 sequences belonging to the *wunen* gene family in *D. melanogaster*.

While other non-*Drosophila* insects have one copy of wunen, lepidopterans had 3 copies (Figure 3.6A). In addition, while other insects had one copy of *wunen-like*, Lepidoptera had 3 copies of *wunen-like* that arose after lepidopteran divergence from other insects (Figure 3.6A). All copies of *wunen* and *wunen-like* were expressed in *M. sexta* heads and in the four tissues of *H. melpomene*. *Wunen-1* and *wunen-like-3* were the two copies expressed more highly in *H. melpomene* heads relative to other tissues suggesting these copies might be the ones having a role in vision.

Loss of a Vha100 Duplicate in Non-Lepidopteran Insects

A gene that was unique to Lepidoptera was *Vha100-1-like* (Figure 3.7A). In *Drosophila*, *Vacuolar H+ ATPase 100kD subunit* 1 (*Vha100-1*) encodes a V_0 sector of a multisubunit complex, vesicular adenosine triphosphatase (v-ATPase) (Williamson et al. 2010). V-ATPase is involved in membrane fusion, acidification and synaptic vesicle exocytosis at photoreceptor presynaptic terminals (Williamson et al. 2010). Loss of *Vha100-1* results in a loss of transient response but overexpression results in photoreceptor cell death (Williamson et al. 2010). In our phylogenetic analysis of *Vha-100*, our results suggests that all insects except Lepidoptera have lost *Vha100-1-like* (Figure 3.7A). *Vha100-1*, *Vha100-1-like*, and *Vha100-2* are all expressed in *M. sexta* heads and in *H. melpomene* heads, antennae, legs, and mouth parts. None of the gene copies are upregulated in heads which suggests that if these genes do have a visual function, that function is not specific and these genes may carry out additional functions in other tissues.

Termination of phototransduction

Homologs of *stops*, *ninaC*, *Arr1*, and *Arr2* were upregulated in *H. melpomene* heads using a transcriptome-wide approach. These genes are necessary in *Drosophila* to deactivate rhodopsin signaling and terminate the phototransduction cascade. *Stops* encodes a protein with a SOCS box (suppressor of cytokine signaling) domain, the *stops* phenotype is associated with slow termination of phototransduction due to a decrease in NORPA (PLC) (Wang et al. 2008). Conversely, the mechanisms of action of Arrestin 1 and Arrestin 2 are to bind light-activated rhodopsin and discontinue cascade signaling (Dolph et al. 1993; Stavenga & Hardie 2011). *NinaC* is also believed to function via arrestin; *ninaC* binds calmodulin, another protein involved in the INAD complex, which accelerates arrestin binding rhodopsin to terminate phototransduction (Venkatachalam et al. 2010). The discovery of these genes upregulated in butterfly heads suggests that phototransduction termination genes might serve a similar function in butterfly vision.

CONCLUSIONS

Previous studies on butterfly and moth vision have focused on classifying the expression of opsin genes and proteins to gain insight into the color vision of these insects. The phototransduction cascade by which the opsins signal light wavelength information to the brain has only been well-characterized in *Drosophila*. Studies of phototransduction in other invertebrates extrapolate from what is known in *Drosophila* to assign potential functions to genes based on sequence similarity. In our study we used transcriptomics and phylogenetics to explore the conservation of phototransduction genes between *Drosophila* and Lepidoptera. We found that many homologs of *Drosophila* phototransduction genes were upregulated in *H. melpomene* heads relative to legs, antennae, and mouth arts. Due to

expression patterns, we suggest a bigger role for TRPL than TRP in lepidopteran vision. Phylogenetic trees for 64 genes showed that many were conserved as single copies between insects, with some exceptions. We found that an unclassified opsin is upregulated in butterfly heads suggesting a role in vision. We identified an instance where a *Drosophila* gene (*inaE*) was lost in Lepidoptera but a paralog of the gene (*DAGL-beta*) may partake in the visual function. Similarly, *wunen* was expanded in Lepidoptera and 2 of the copies, *wunen1* and *wunen-lik3*, were upregulated in butterfly heads suggesting a role in vision. We also found a gene, *Vha100-1-like*, that is lost is non-lepidopteran insects. Our results suggest that the phototransduction cascade is conserved between Lepidoptera and *Drosophila* except for instances where Lepidopteran-specific copies take on a role in vision.

REFERENCES

Arendt D. 2004. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. Science. 306:869–871. doi: 10.1126/science.1099955.

Arendt D. 2003. Evolution of eyes and photoreceptor cell types. Int. J. Dev. Biol. 47:563–571. doi: 10.1387/IJDB.14756332.

Bähner M, Sander P, Paulsen R, Huber A. 2000. The visual G protein of fly photoreceptors interacts with the PDZ domain assembled INAD signaling complex via direct binding of activated Gα(q) to phospholipase Cβ. J. Biol. Chem. 275:2901–2904. doi: 10.1074/jbc.275.4.2901.

Bao R, Friedrich M. 2009. Molecular evolution of the *Drosophila* retinome: exceptional gene gain in the higher Diptera. Mol. Biol. Evol. 26:1273–87. doi: 10.1093/molbev/msp039.

Battelle BA et al. 2016. Opsin repertoire and expression patterns in horseshoe crabs: evidence from the genome of *Limulus polyphemus* (Arthropoda: Chelicerata). Genome Biol. Evol. 8:1571–1589. doi: 10.1093/gbe/evw100.

Bloomquist BT et al. 1988. Isolation of a putative phospholipase C gene of *Drosophila*, *norpA*, and its role in phototransduction. Cell. 54:723–733. doi: 10.1016/S0092-8674(88)80017-5.

Briscoe AD et al. 2010. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. Proc. Natl. Acad. Sci. U. S. A. 107:3628–33. doi: 10.1073/pnas.0910085107.

Brotz TM, Gundelfinger ED, Borst A. 2001. Cholinergic and GABAergic pathways in fly motion vision. BMC Neurosci. 2:1. doi: 10.1186/1471-2202-2-1.

Camacho C et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10:421. doi: Artn 421\nDoi 10.1186/1471-2105-10-421.

Chevesich J, Kreuz AJ, Montell C. 1997. Requirement for the PDZ domain protein, INAD, for localization of the TRP store-operated channel to a signaling complex. Neuron. 18:95–105. doi: 10.1016/S0896-6273(01)80049-0.

Chung BY, Kilman VL, Keath JR, Pitman JL, Allada R. 2009. The GABAA receptor RDL acts in peptidergic PDF neurons to promote sleep in *Drosophila*. Curr. Biol. 19:386–390. doi: 10.1016/j.cub.2009.01.040.

Chyb S, Raghu P, Hardie RC. 1999. Polyunsaturated fatty acids activate the *Drosophila* light-sensitive channels TRP and TRPL. Nature. 397:255–259. doi: 10.1038/16703.

Conesa A et al. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674–3676. doi: 10.1093/bioinformatics/bti610.

Conesa A, Götz S. 2008. Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int. J. Plant Genomics. 619832. doi: 10.1155/2008/619832.

Davey JW et al. 2016. Major improvements to the *Heliconius melpomene* genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. G3. 6:695–708.

Dolph PJ et al. 1993. Arrestin function in inactivation of G protein-coupled receptor rhodopsin in vivo. Science. 260:1910–1916. doi: 10.1126/science.8316831.

Duan Y et al. 2017. Transcriptome analysis of molecular mechanisms responsible for light-stress response in *Mythimna separata* (Walker). Sci. Rep. 7:45188. doi: 10.1038/srep45188.

Edgar RC. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797. doi: 10.1093/nar/gkh340.

Faggionato D, Serb JM. 2017. Strategy to identify and test putative light-sensitive non-opsin G-protein-coupled receptors: A case study. Biol. Bull. 233:70–82. doi: 10.1086/694842.

Fain GL, Hardie R, Laughlin SB. 2010. Phototransduction and the evolution of photoreceptors. Curr. Biol. 20:R114-24. doi: 10.1016/j.cub.2009.12.006.

Feuda R, Marle F, Bentley MA, Holland PWH. 2016. Conservation, duplication, and divergence of five opsin genes in insect evolution. Genome Biol. Evol. 8:579–587. doi: 10.5287/bod-leian.

Fichelson P, Brigui A, Pichaud F. 2012. Orthodenticle and Kruppel homolog 1 regulate *Drosophila* photoreceptor maturation. Proc. Natl. Acad. Sci. U. S. A. 109:7893–8. doi: 10.1073/pnas.1120276109.

French AS, Meisner S, Liu H, Weckström M, Torkkeli PH. 2015. Transcriptome analysis and RNA interference of cockroach phototransduction indicate three opsins and suggest a major role for TRPL channels. Front. Physiol. 6:207. doi: 10.3389/fphys.2015.00207.

Friedrich M et al. 2011. Phototransduction and clock gene expression in the troglobiont beetle *Ptomaphagus hirtus* of Mammoth cave. J. Exp. Biol. 214:3532–41.

Futahashi R et al. 2015. Extraordinary diversity of visual opsin genes in dragonflies. Proc. Natl. Acad. Sci. 112:E1247–E1256. doi: 10.1073/pnas.1424670112.

Garcia-Murillas I et al. 2006. *lazaro* encodes a lipid phosphate phosphohydrolase that regulates phosphatidylinositol turnover during *Drosophila* phototransduction. Neuron. 49:533–546. doi: 10.1016/j.neuron.2006.02.001.

Gengs C et al. 2002. The target of *Drosophila* photoreceptor synaptic transmission is a histamine-gated chloride channel encoded by *ort* (*hclA*). J. Biol. Chem. 277:42113–42120. doi: 10.1074/jbc.M207133200.

Giraldo-Calderón GI, Zanis MJ, Hill CA. 2017. Retention of duplicated long-wavelength opsins in mosquito lineages by positive selection and differential expression. BMC Evol. Biol. 17:84. doi: 10.1186/s12862-017-0910-6.

Girardot F, Lasbleiz C, Monnier V, Tricoire H. 2006. Specific age related signatures in *Drosophila* body parts transcriptome. BMC Genomics. 7:69. doi: 10.1186/1471-2164-7-69.

Götz S et al. 2011. B2G-FAR, a species-centered GO annotation repository. Bioinformatics. 27:919–924. doi: 10.1093/bioinformatics/btr059.

Götz S et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420–3435. doi: 10.1093/nar/gkn176.

Hardie RC. 2001. Phototransduction in *Drosophila melanogaster*. J. Exp. Biol. 204:3403–3409.

Hardie RC, Juusola M. 2015. Phototransduction in *Drosophila*. Curr. Opin. Neurobiol. 34:37–45. doi: 10.1016/j.conb.2015.01.008.

Hardie RC, Minke B. 1992. The *trp* gene is essential for a light-activated Ca2+ channel in *Drosophila* photoreceptors. Neuron. 8:643–651. doi: 10.1016/0896-6273(92)90086-S.

Hardie RC, Raghu P. 2001. Visual transduction in *Drosophila*. Nature. 413:186–193. doi: 10.1038/35093002.

Henze MJ, Dannenhauer K, Kohler M, Labhart T, Gesemann M. 2012. Opsin evolution and expression in arthropod compound eyes and ocelli : Insights from the cricket *Gryllus bimaculatus*. BMC Evol. Biol. 12:163.

Horridge GA, Giddings C, Stange G. 1972. The superposition eye of skipper butterflies. Proc. R. Soc. London. Ser. B, Biol. Sci. 182:457–495. doi: 10.1098/rspb.1972.0088.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4:44–57. doi: 10.1038/nprot.2008.211.

Ile KE, Tripathy R, Goldfinger V, Renault AD. 2012. Wunen, a *Drosophila* lipid phosphate phosphatase, is required for septate junction-mediated barrier function. Development. 139:2535–2546. doi: 10.1242/dev.077289.

Kanost MR et al. 2016. Multifaceted biological insights from a draft genome sequence of the tobacco hornworm moth, *Manduca sexta*. Insect Biochem. Mol. Biol. 76:118–147. doi: 10.1016/j.ibmb.2016.07.005.

Katz B, Minke B. 2009. *Drosophila* photoreceptors and signaling mechanisms. Front. Cell. Neurosci. 3:2. doi: 10.3389/neuro.03.002.2009.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. Mol. Biol. Evol. 33:1870–1874. doi: 10.1093/molbev/msw054.

Lee YJ et al. 1994. The *Drosophila dgq* gene encodes a G alpha protein that mediates phototransduction. Neuron. 13:1143–57. doi: 10.1016/0896-6273(94)90052-3.

Leung H et al. 2008. DAG lipase activity is necessary for TRP channel regulation in *Drosophila* photoreceptors. Neuron. 58:884–896.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323. doi: 10.1186/1471-2105-12-323.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25:1754–1760. doi: 10.1093/bioinformatics/btp324.

von Lintig J, Kiser PD, Golczak M, Palczewski K. 2010. The biochemical and structural basis for *trans*-to-*cis* isomerization of retinoids in the chemistry of vision. Trends Biochem. Sci. 35:400–410. doi: 10.1016/j.tibs.2010.01.005.

Liu X, Krause WC, Davis RL. 2007. GABAA receptor RDL inhibits *Drosophila* olfactory associative learning. Neuron. 56:1090–1102. doi: 10.1016/j.neuron.2007.10.036.

Macias-Muñoz A, McCulloch KJ, Briscoe AD. 2017. Copy number variation and expression analysis reveals a non-orthologous *pinta* gene family member involved in butterfly vision. Genome Biol. Evol. 9:3398–3412. doi: 10.1093/gbe/evx230.

Macias-Muñoz A, Smith G, Monteiro A, Briscoe AD. 2016. Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic. Mol. Biol. Evol. 33:79–92. doi: 10.1093/molbev/msv197.

Mahato S et al. 2014. Common transcriptional mechanisms for visual photoreceptor cell differentiation among pancrustaceans. PLoS Genet. 10:e1004484. doi: 10.1371/journal.pgen.1004484.

Marygold SJ et al. 2013. FlyBase: Improvements to the bibliography. Nucleic Acids Res. 41:D751–D757. doi: 10.1093/nar/gks1024.

McCulloch KJ et al. 2017. Sexual dimorphism and retinal mosaic diversification following the evolution of a violet receptor in butterflies. Mol. Biol. Evol. 34:2271–2284. doi: 10.1093/molbev/msx163.

McCulloch KJ, Osorio D, Briscoe AD. 2016. Sexual dimorphism in the compound eye of *Heliconius erato*: a nymphalid butterfly with at least five spectral classes of photoreceptor. J. Exp. Biol. 219:2377–2387. doi: 10.1242/jeb.136523.

Min XJ, Butler G, Storms R, Tsang A. 2005. OrfPredictor: Predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33:677–680. doi: 10.1093/nar/gki394.

Montell C. 2012. *Drosophila* visual transduction. Trends Neurosci. 35:356–363. doi: 10.1016/j.tins.2012.03.004.Drosophila.

Montell C. 2005. TRP channels in *Drosophila* photoreceptor cells. J. Physiol. 567:45–51. doi: 10.1113/jphysiol.2005.092551.

Montell C, Birnbaumer L, Flockerzi V. 2002. The TRP channels, a remarkably functional family. Cell. 108:595–598. doi: 10.1016/S0092-8674(02)00670-0.

Montell C, Rubin GM. 1989. Molecular characterization of the *Drosophila trp* locus: a putative integral membrane protein required for phototransduction. Neuron. 2:1313–1323. doi: 10.1016/0896-6273(89)90069-X.

Montgomery SH, Merrill RM, Ott SR. 2016. Brain composition in *Heliconius* butterflies, posteclosion growth and experience-dependent neuropil plasticity. J. Comp. Neurol. 524:1747–1769. doi: 10.1002/cne.23993.

Ni JD, Baik LS, Holmes TC, Montell C. 2017. A rhodopsin in the brain functions in circadian photoentrainment in *Drosophila*. Nature. 545:340–344. doi: 10.1038/nature22325.

Niemeyer BA, Suzuki E, Scott K, Jalink K, Zuker CS. 1996. The *Drosophila* light-activated conductance is composed of the two channels TRP and TRPL. Cell. 85:651–659. doi: 10.1016/S0092-8674(00)81232-5.

Nilsson DE, Land MF, Howard J. 1984. Afocal apposition optics in butterfly eyes. Nature. 312:561–563. doi: 10.1038/312561a0.

Perry M et al. 2016. Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. Nature. 535:280–4. doi: 10.1038/nature18616.

Plachetzki DC, Degnan BM, Oakley TH. 2007. The origins of novel protein interactions during animal opsin evolution. PLoS One. 2. doi: 10.1371/journal.pone.0001054.

Plachetzki DC, Fong CR, Oakley TH. 2010. The evolution of phototransduction from an ancestral cyclic nucleotide gated pathway. Proc. R. Soc. B Biol. Sci. 277:1963–1969. doi: 10.1098/rspb.2009.1797.

Ploner A. 2012. Heatplus: Heatmaps with row and/or column covariates and colored clusters.

Porter ML et al. 2012. Shedding new light on opsin evolution. Proc. R. Soc. B Biol. Sci. 279:3–14. doi: 10.1098/rspb.2011.1819.

Raible F et al. 2006. Opsins and clusters of sensory G-protein-coupled receptors in the sea urchin genome. Dev. Biol. 300:461–475. doi: 10.1016/j.ydbio.2006.08.070.

Ramirez MD et al. 2016. The last common ancestor of most bilaterian animals possessed at least nine opsins. Genome Biol. Evol. 8:3640–3652. doi: 10.1093/gbe/evw248.

Rivera AS et al. 2010. Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. BMC Evol. Biol. 10:123. doi: 10.1186/1471-2148-10-123.

Robinson M, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25. doi: 10.1186/gb-2010-11-3-r25.

Robinson MD, Mccarthy DJ, Smyth GK. 2010. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. 26:139–140. doi: 10.1093/bioinformatics/btp616.

Schwarz G. 1978. Estimating the dimension of a model. Ann. Stat. 6:461–464.

Seymoure BM, Mcmillan WO, Rutowski R. 2015. Peripheral eye dimensions in Longwing (*Heliconius*) butterflies vary with body size and sex but not light environment nor mimicry ring. J. Res. Lepid. 48:83–92.

Shichida Y, Matsuyama T. 2009. Evolution of opsins and phototransduction. Philos. Trans. R. Soc. B Biol. Sci. 364:2881–2895. doi: 10.1098/rstb.2009.0051.

Shieh BH, Stamnes MA, Seavello S, Harris GL, Zuker CS. 1989. The *ninaA* gene required for visual transduction in *Drosophila* encodes a homologue of cyclosporin A-binding protein. Nature. 338:67–70. doi: 10.1038/338067a0.

Shieh BH, Zhu MY. 1996. Regulation of the TRP Ca2+ channel by INAD in drosophila photoreceptors. Neuron. 16:991–998. doi: 10.1016/S0896-6273(00)80122-1.

Smith G, Briscoe AD. 2015. Molecular evolution and expression of the CRAL_TRIO protein family in insects. Insect Biochem. Mol. Biol. 62:168–173. doi: 10.1016/j.ibmb.2015.02.003.

Smith G, Chen YR, Blissard GW, Briscoe AD. 2014. Complete dosage compensation and sex-biased gene expression in the moth *Manduca sexta*. Genome Biol. Evol. 6:526–537. doi: 10.1093/gbe/evu035.

Spaethe J, Briscoe AD. 2004. Early duplication and functional diversification of the opsin gene family in insects. Mol. Biol. Evol. 21:1583–1594. doi: 10.1093/molbev/msh162.

Spaethe J, Briscoe AD. 2005. Molecular characterization and expression of the UV opsin in bumblebees : three ommatidial subtypes in the retina and a new photoreceptor organ in the lamina. 2347–2361. doi: 10.1242/jeb.01634.

Stavenga DG, Hardie RC. 2011. Metarhodopsin control by arrestin, light-filtering screening pigments, and visual pigment turnover in invertebrate microvillar photoreceptors. J. Comp. Physiol. A Neuroethol. Sensory, Neural, Behav. Physiol. 197:227–241. doi: 10.1007/s00359-010-0604-7.

Suga H, Schmid V, Gehring WJ. 2008. Evolution and functional diversity of jellyfish opsins. Curr. Biol. 18:51–55. doi: 10.1016/j.cub.2007.11.059.

Tsunoda S et al. 1997. A multivalent PDZ-domain protein assembles signalling complexes in a G- protein-coupled cascade. Nature. 388:243–249. doi: 10.1038/40805.

Venkatachalam K et al. 2010. Dependence on a retinophilin/myosin complex for stability of PKC and INAD and termination of phototransduction. J. Neurosci. 30:11337–11345. doi: 10.1523/JNEUROSCI.2709-10.2010.Dependence.

Vöcking O, Kourtesis I, Tumu S, Hausen H. 2017. Co-expression of xenopsin and rhabdomeric opsin in photoreceptors bearing microvilli and cilia. Elife. 6:e23435. doi: 10.7554/eLife.23435.

Wang T et al. 2005. Light activation, adaptation, and cell survival functions of the Na +/Ca2+ exchanger CalX. Neuron. 45:367–378. doi: 10.1016/j.neuron.2004.12.046.

Wang T, Montell C. 2005. Rhodopsin formation in *Drosophila* is dependent on the PINTA retinoid-binding protein. J. Neurosci. 25:5187–94.

Wang T, Wang X, Xie Q, Montell C. 2008. The SOCS box protein STOPS is required for phototransduction through its effects on phospholipase C. Neuron. 57:56–68. doi: 10.1016/j.neuron.2007.11.020.

Warrant E, Dacke M. 2016. Visual navigation in nocturnal insects. Physiology. 31:182–192. doi: 10.1152/physiol.00046.2015.

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Williamson WR, Wang D, Haberman AS, Hiesinger PR. 2010. A dual function of V0-ATPase a1 provides an endolysosomal degradation mechanism in *Drosophila melanogaster* photoreceptors. J. Cell Biol. 189:885–899. doi: 10.1083/jcb.201003062.

Yack JE, Johnson SE, Brown SG, Warrant EJ. 2007. The eyes of *Macrosoma* sp. (Lepidoptera: Hedyloidea): A nocturnal butterfly with superposition optics. Arthropod Struct. Dev. 36:11–22. doi: 10.1016/j.asd.2006.07.001.

Yagi N, Koyama N. 1963. The compound eye of Lepidoptera: Approach from organic evolution. In: Shinkyo-Press.

Yuan F, Bernard GD, Le J, Briscoe AD. 2010. Contrasting modes of evolution of the visual pigments in *Heliconius* butterflies. Mol. Biol. Evol. 27:2392–2405. doi: 10.1093/molbev/msq124.

Yuan Q, Song Y, Yang C-H, Jan LY, Jan YN. 2013. Female contact modulates male aggression via a sexually dimorphic GABAergic circuit in *Drosophila*. Nat. Neurosci. 17:81–88. doi: 10.1038/nn.3581.

Zhang N, Zhang J, Purcell KJ, Cheng Y, Howard K. 1997. The *Drosophila* protein wunen repels migrating germ cells. Nature. 385:64–67. doi: 10.1038/385064a0.

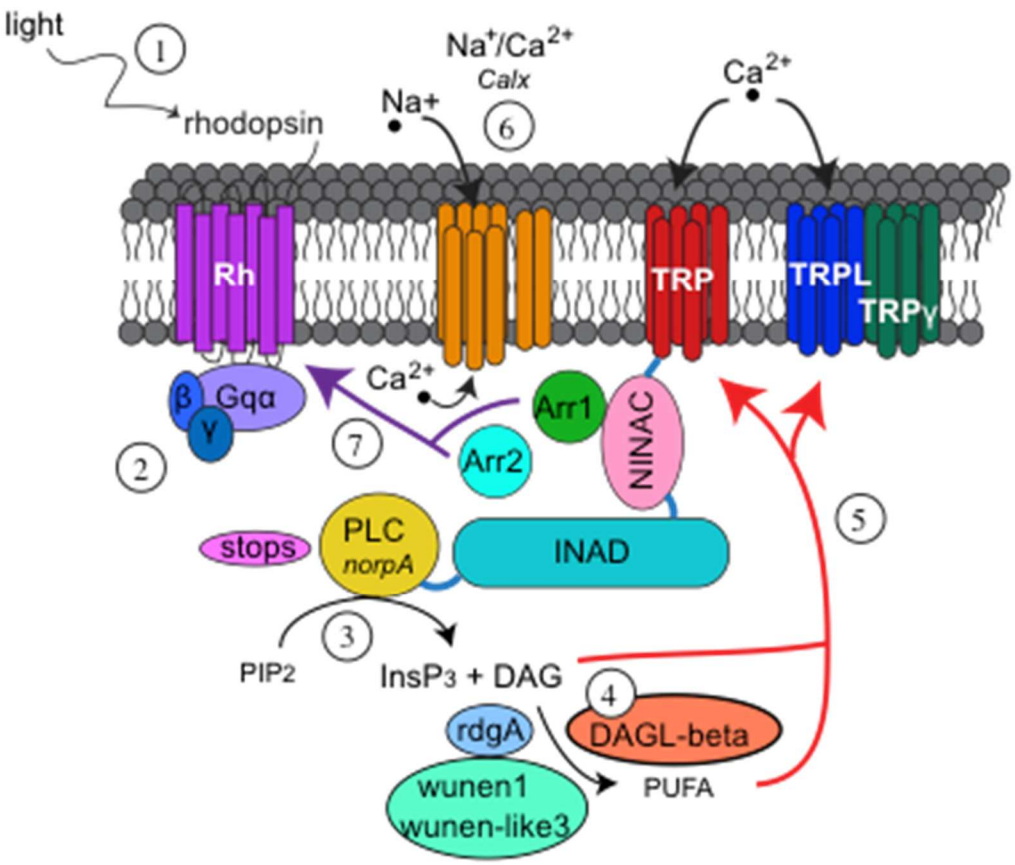
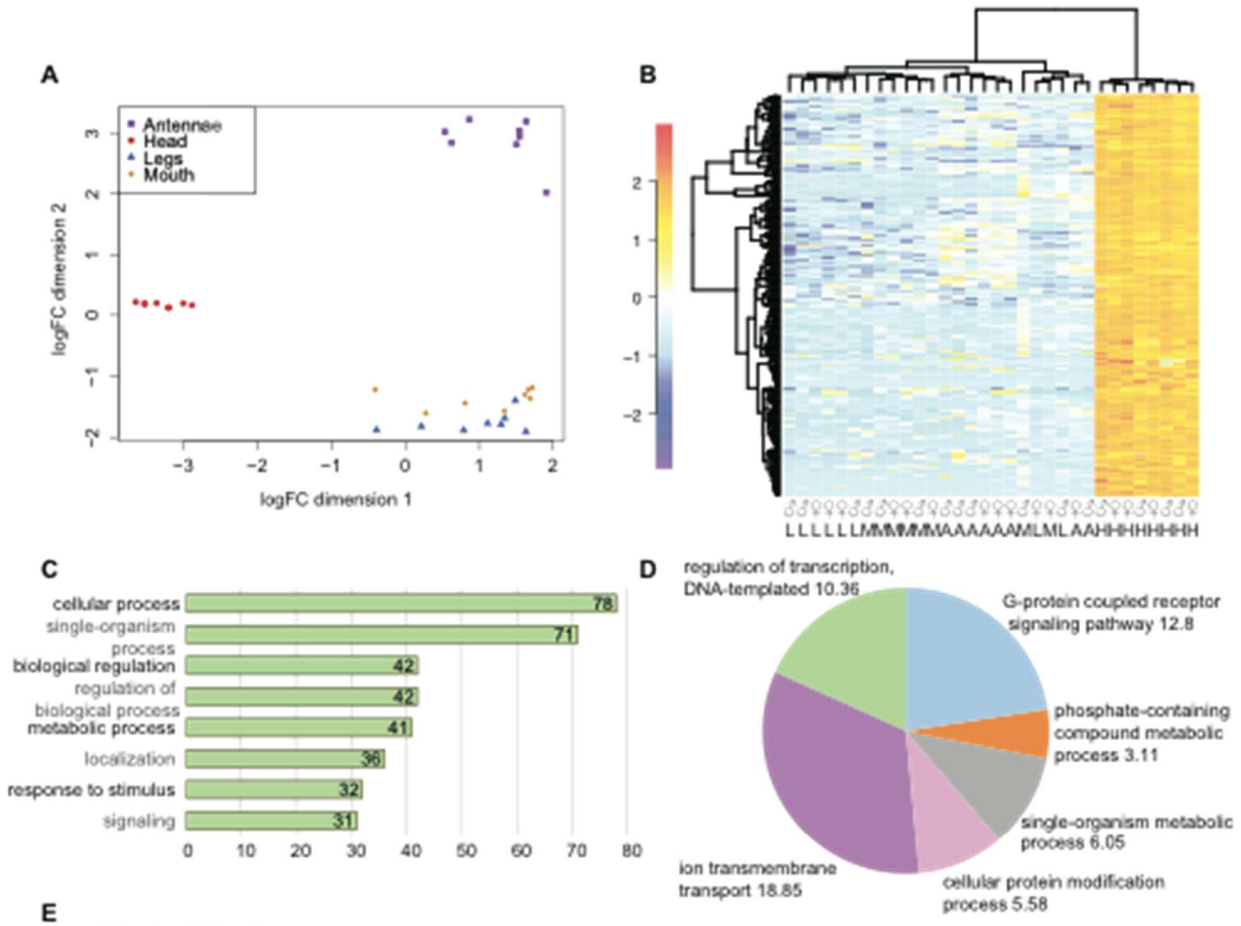


Figure 3.1. Proposed model of the lepidopteran phototransduction cascade. 1) Light activates rhodopsin by a configurational change of the chromophore from 11-*cis* to all-*trans*. 2) $G_q\alpha$ is released from a G-protein complex of 3 subunits (α, β, and γ) and activates phospholipase C (PLC). 3) PLC hydrolyses phosphatidylinositol 4,5-bisphosphate (PIP_2) to produce inositol 1,4,5-trisphosphate ($InsP_3$) and diacylglycerol (DAG). 4) Diacylglycerol lipase (DAGL-beta) hydrolyzes DAG to produce polyunsaturated fatty acids (PUFAs). 5) DAG and PUFA may activate TRP and TRPL, mechanism has not been established. 6) Na^+/Ca^{2+} exchanger channel pumps Ca^{2+} out of the photoreceptor cell. 7) Arrestin 1 and 2 bind rhodopsin to terminate the cascade.



E

Annotation Cluster	Enrichment Score	Term	Count	Genes	PValue	Bonferroni
1	9.757	detection of light stimulus	12	NINAC, INAD, CALX, ARR2, ARR1, RH3, RH5, STOPS, TRPL, NORPA, GBETA76C, NINAA	1.85E-12	1.06E-09
2	8.134	regulation of rhodopsin mediated signaling pathway	7	NINAC, INAD, ARR2, ARR1, STOPS, NORPA, GBETA76C	2.44E-09	1.40E-06
3	8.077	detection of light stimulus involved in visual perception	8	NINAC, INAD, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	8.40E-10	4.81E-07
4	6.062	ligand-gated ion channel activity	11	RDL, CG8916, NMDAR1, LCCH3, HISCL1, ORT, TRPL, CG42260, CG3822, CG11155, NMDAR2	2.47E-09	5.83E-07
5	3.739	Gamma-aminobutyric acid A receptor	5	RDL, CG8916, LCCH3, HISCL1, ORT	5.94E-06	0.001
6	3.085	synaptic transmission	9	IPP, SYT1, NMDAR1, CPX, ORT, CG1909, GAD1, VGLUT, SIF	5.97E-04	0.290
7	2.715	NMDA receptor	4	NMDAR1, CG11155, NMDAR2, CG3822	3.23E-04	0.069
8	2.652	chloride channel complex	4	RDL, LCCH3, HISCL1, ORT	4.90E-04	0.049
9	2.307	Homeobox	6	CG11294, BSH, HMX, PPH13, SCRO, OC	0.002	0.287
10	1.288	rhabdomere development	4	PPH13, OC, PROM, CHP	0.001	0.532
11	1.144	C2 membrane targeting protein	3	SYT1, SL, NORPA	0.046	1.000

Figure 3.2. Differential expression analysis. (A) Multidimensional scaling (MDS) plot of RNA-seq libraries from H. melpomene antennae, head, legs and mouth parts. (B) Heatmap of genes commonly upregulated in heads, numbers indicate log-fold change. (C) Level 2 biological process terms for genes commonly upregulated in heads using Blast2GO. (D) Multilevel pie chart summary of GO terms with node score information using Blast2GO. A node score is the number of sequences associated to a particular GO term. (E) Enrichment results for genes commonly upregulated in heads and annotated using FlyBase and DAVID.

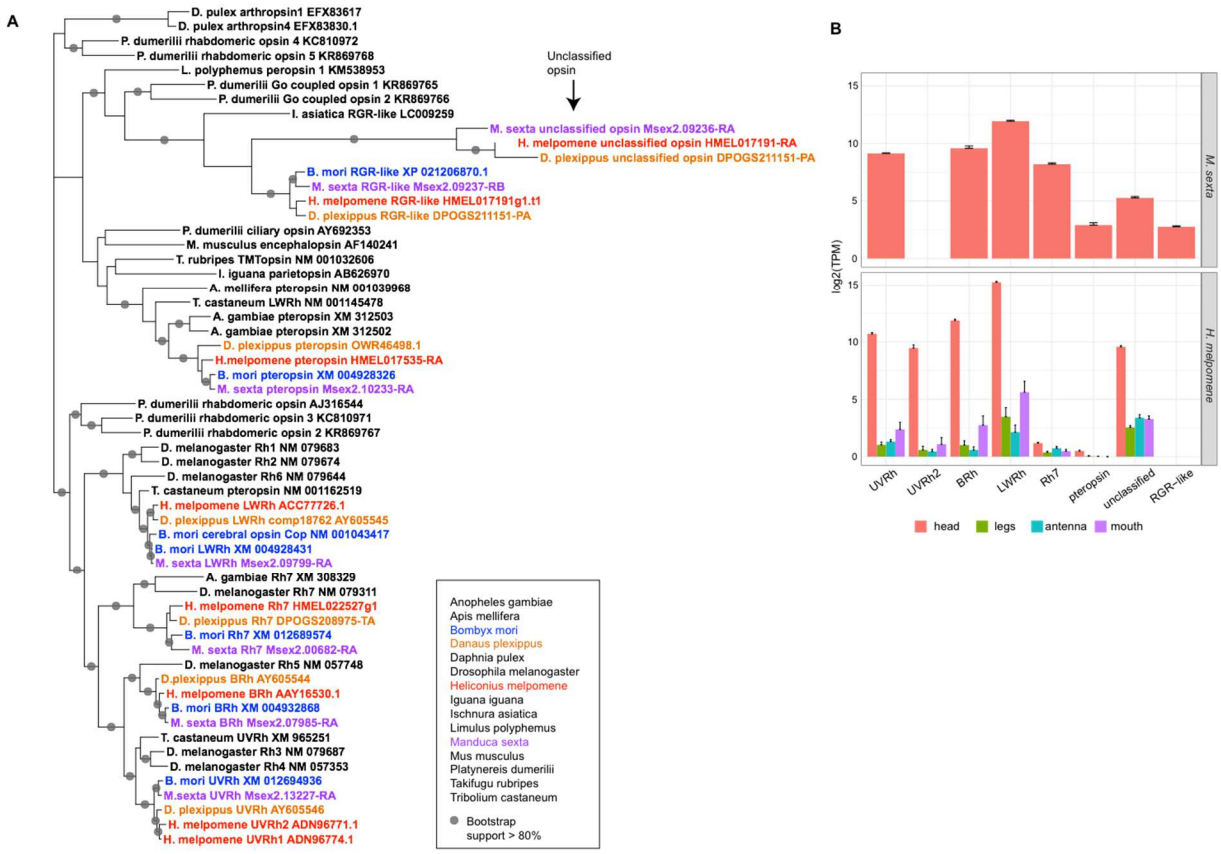


Figure 3.3. Opsin gene family. (A) *Opsin* phylogenetic tree generated using sequences from Kanost et al. (2016) and sequences from *H. melpomene* and *D. plexippus*. (B) Expression of opsin genes in *M. sexta* heads (n=8) and *H. melpomene* heads (n=8), antennae (n=8), legs (n=8), and mouth parts (n=8).

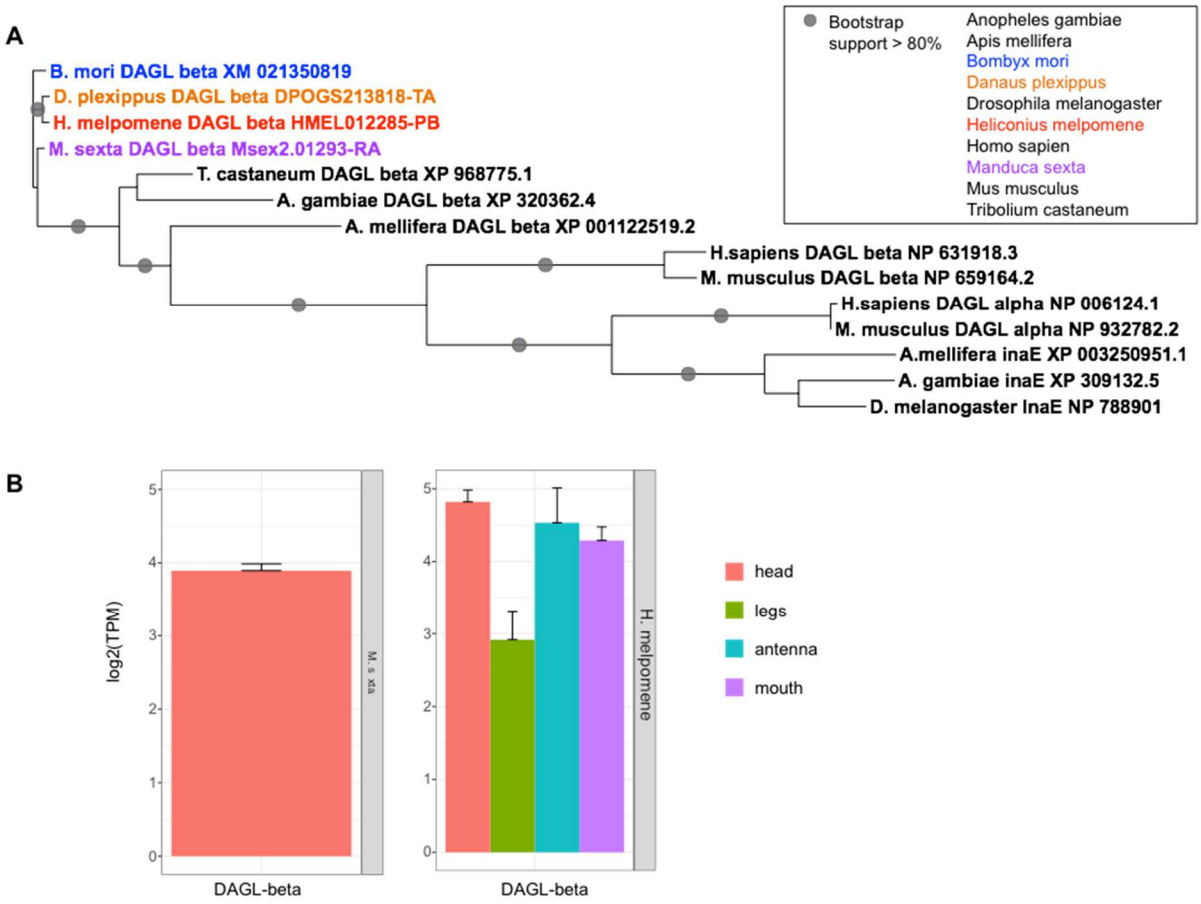


Figure 3.4. DAGL gene family. (A) *DAGL* phylogenetic tree generated using sequences from 8 insect genome and *H. sapien* and *M. musculus.* (B) Expression of DAGL genes in *M. sexta* heads (n=8) and *H. melpomene* heads (n=8), antennae (n=8), legs (n=8), and mouth parts (n=8).

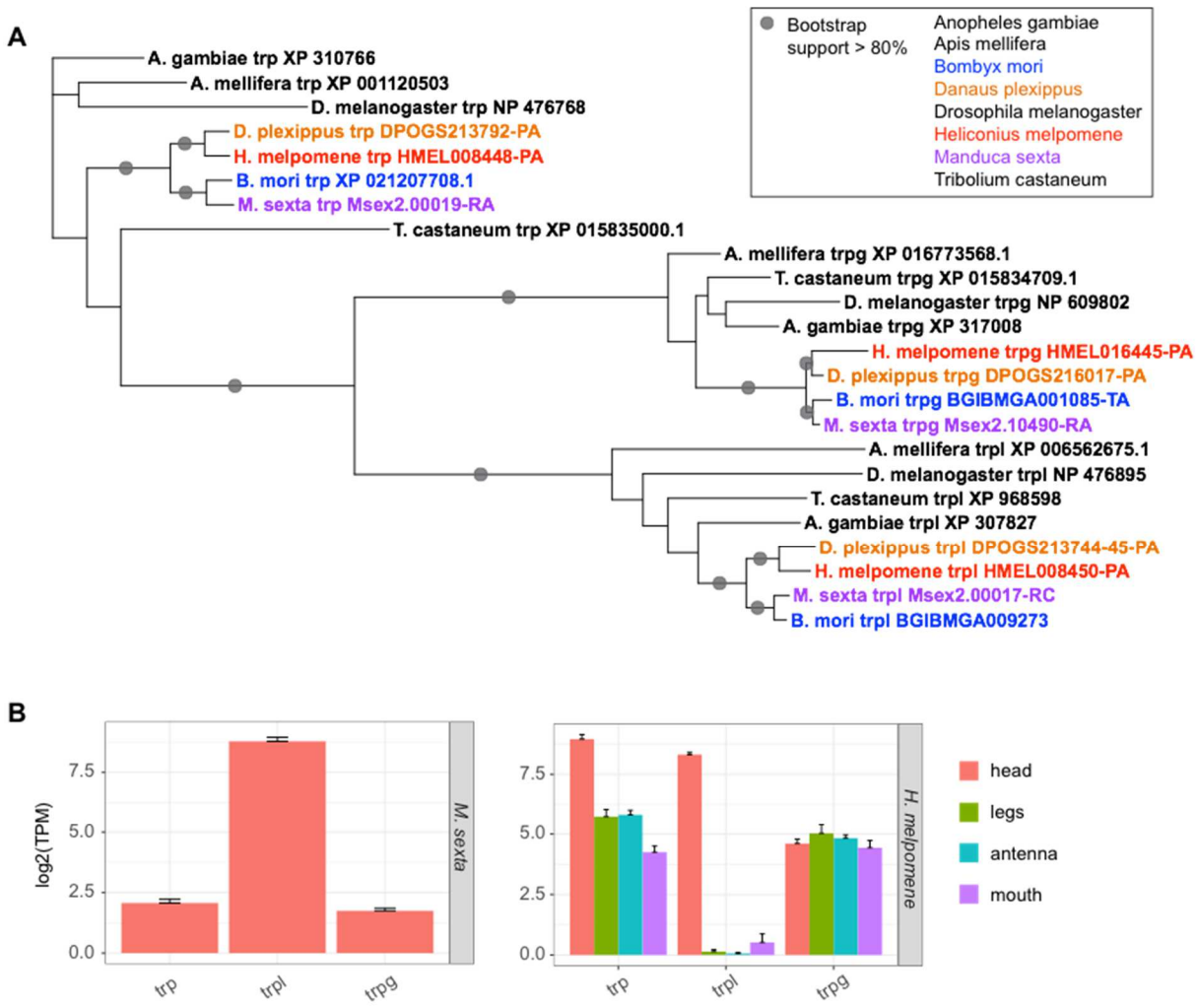


Figure 3.5. TRP gene family. (A) *TRP* phylogenetic tree generated using sequences from 8 insect genomes. (B) Expression of TRP genes in *M. sexta* heads (n=8) and *H. melpomene* heads (n=8), antennae (n=8), legs (n=8), and mouth parts (n=8).

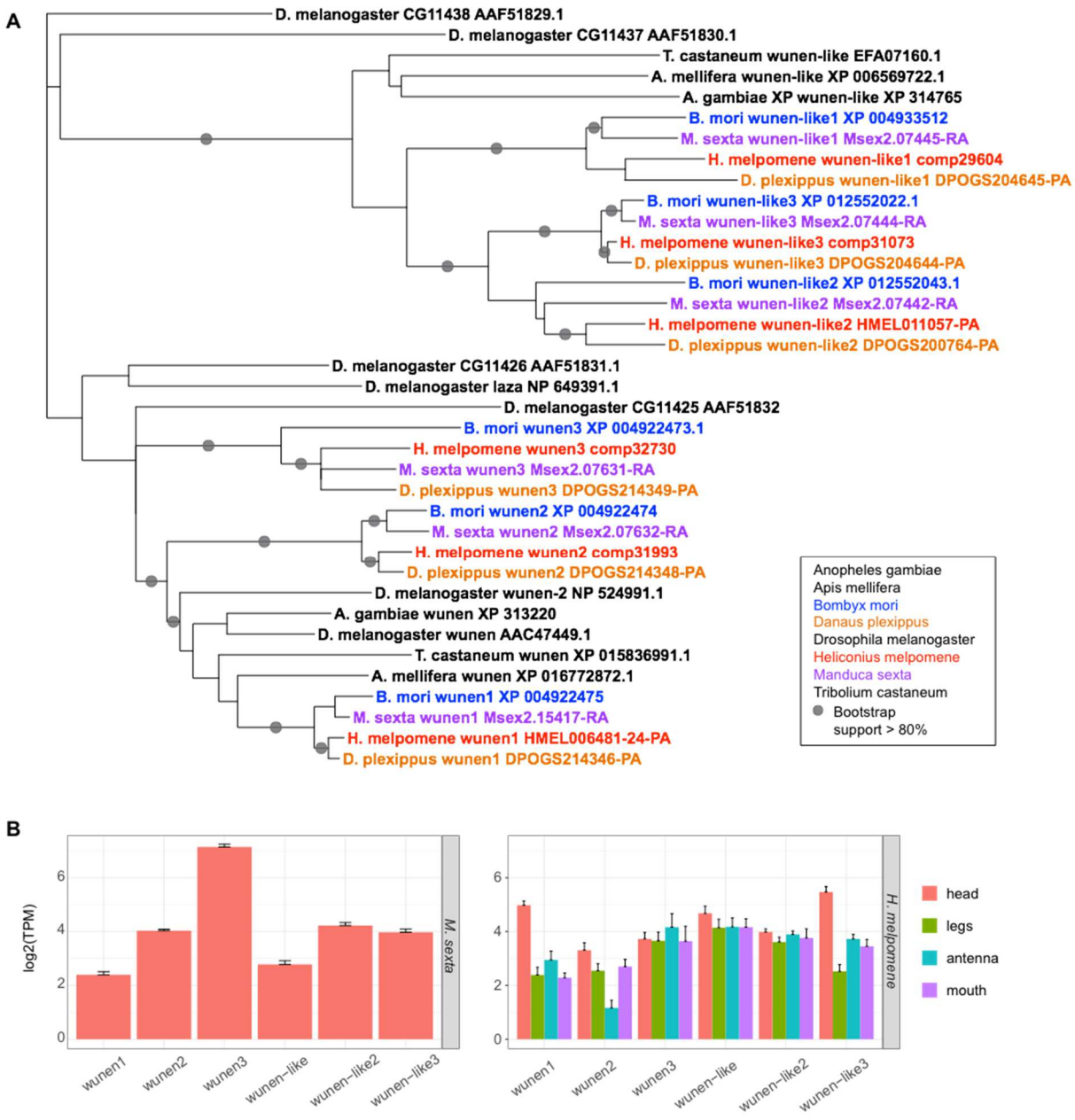


Figure 3.6. Wunen gene family. (A) *Wunen* phylogenetic tree generated using sequences from 8 insect genomes. (B) Expression of wunen genes in *M. sexta* heads (n=8) and *H. melpomene* heads (n=8), antennae (n=8), legs (n=8), and mouth parts (n=8).

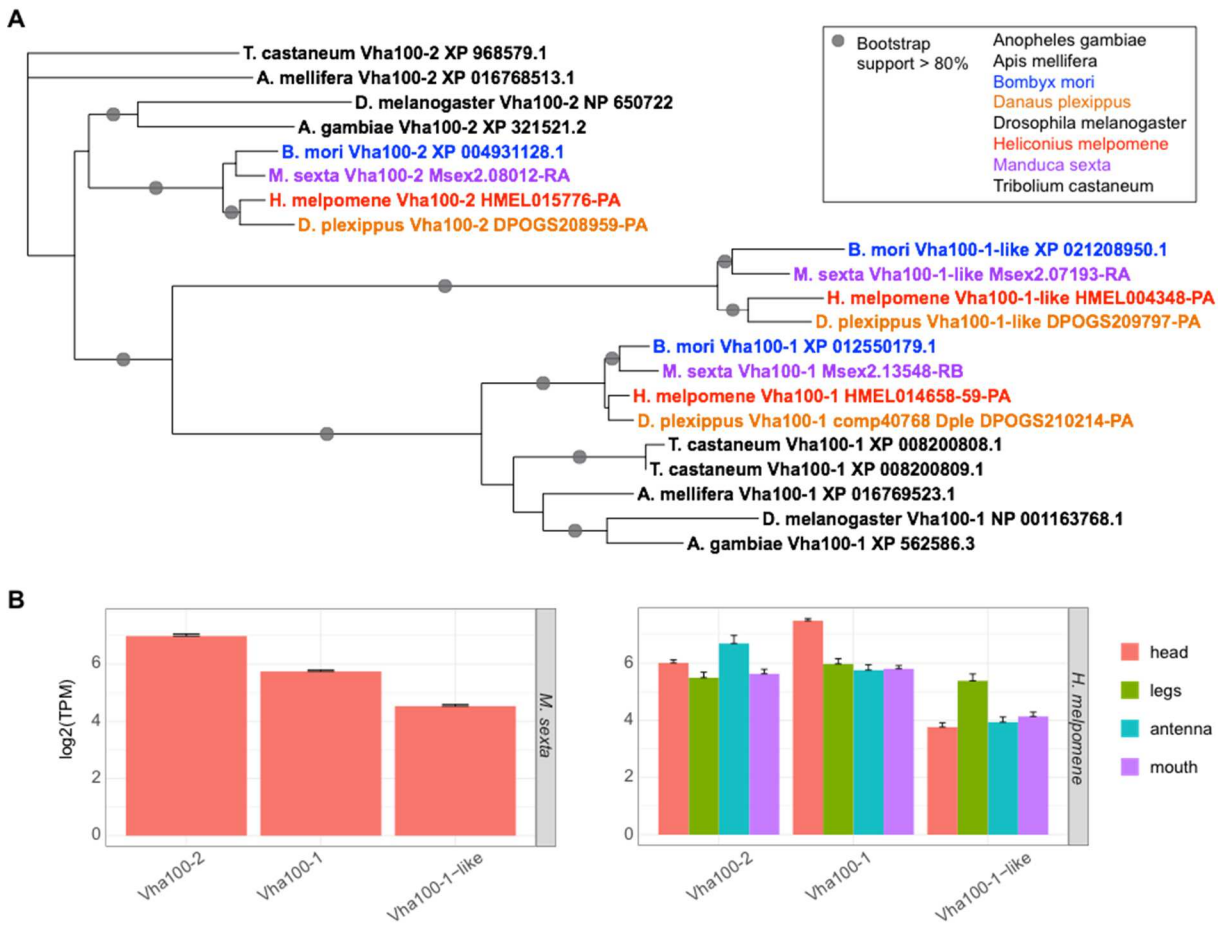


Figure 3.7. Vha-100 gene family. (A) *Vha-100* phylogenetic tree generated using sequences from 8 insect genomes. (B) Expression of Vha-100 genes in in *M. sexta* heads (n=8) and *H. melpomene* heads (n=8), antennae (n=8), legs (n=8), and mouth parts (n=8).

Table 3.1. Summary of transcriptome-wide analysis.

	qvalue	Bonferroni	Upregulated in Heads*	Commonly Upregulated in heads	Unique GO terms
Head vs. Antennae	4,868	1,173	561		590
Head vs. Legs	6,108	1,472	928		748
Head vs. Mouth	6,176	1,486	914		700
Merged				281	154

*Merged are genes commonly upregulated in heads after merging results of pairwise comparisons.

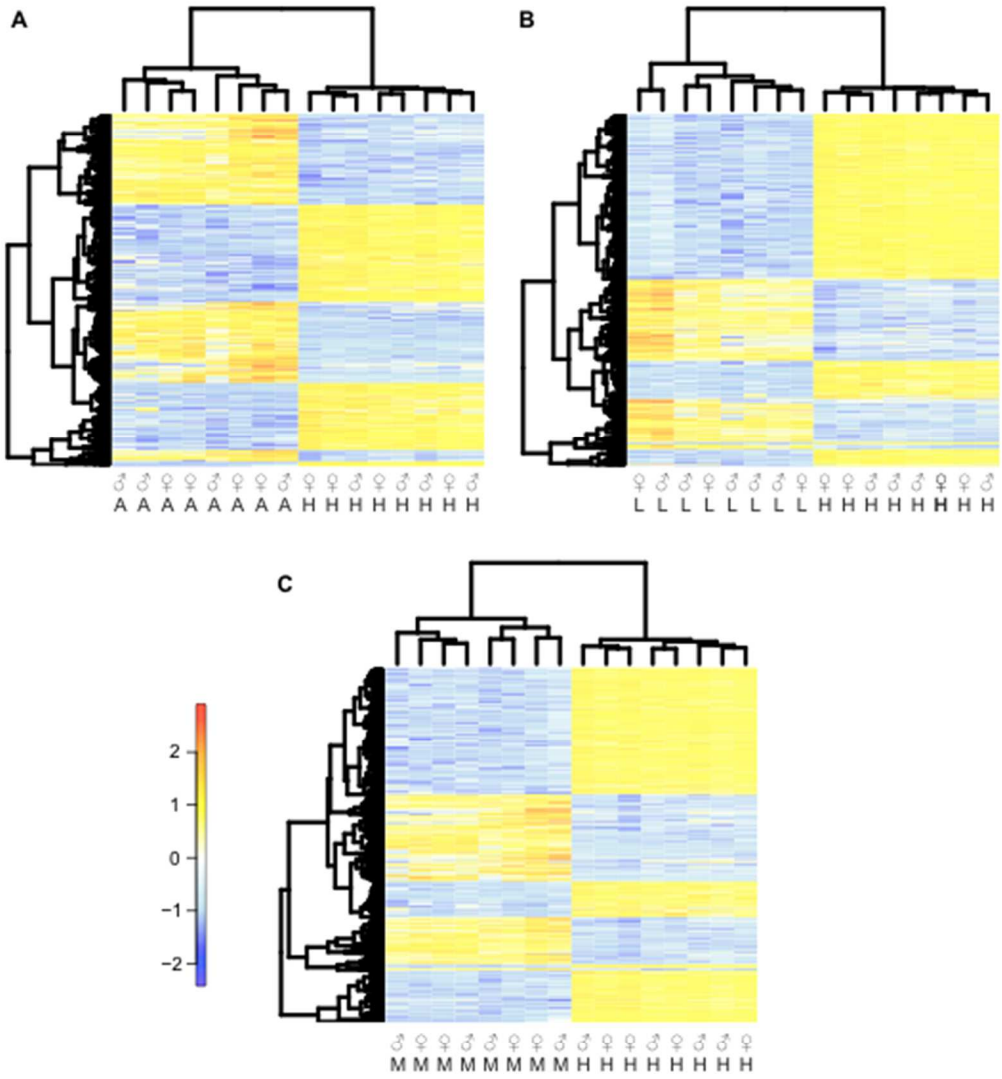


Figure S3.1. Heatmaps of head vs. antennae, legs and mouthparts. (A) Head versus antennae. (B) Head versus legs. (C) Head versus mouth parts.

Table S3.1. Accession numbers for RNA-Seq data used

Species	Specimen	Sex	Tissue	Accession
Manduca Sexta	female_head_1	F	head	E-MTAB-2066
Manduca Sexta	female_head_2	F	head	E-MTAB-2066
Manduca Sexta	female_head_3	F	head	E-MTAB-2066
Manduca Sexta	female_head_4	F	head	E-MTAB-2066
Manduca Sexta	male_head_1	M	head	E-MTAB-2066
Manduca Sexta	male_head_2	M	head	E-MTAB-2066
Manduca Sexta	male_head_3	M	head	E-MTAB-2066
Manduca Sexta	male_head_4	M	head	E-MTAB-2066
Heliconius melpomene	HMP110	M	head	E-MTAB-6342
Heliconius melpomene	HMP112	F	head	E-MTAB-6342
Heliconius melpomene	HMP114	M	head	E-MTAB-6342
Heliconius melpomene	HMP115	F	head	E-MTAB-6342
Heliconius melpomene	HMP333	M	head	E-MTAB-6249
Heliconius melpomene	HMP347	F	ant	E-MTAB-1500
Heliconius melpomene	HMP347	F	head	E-MTAB-6249
Heliconius melpomene	HMP347	F	legs	E-MTAB-1500
Heliconius melpomene	HMP347	F	mou	E-MTAB-1500
Heliconius melpomene	HMP406	M	ant	E-MTAB-1500
Heliconius melpomene	HMP406	M	legs	E-MTAB-1500
Heliconius melpomene	HMP406	M	mou	E-MTAB-1500
Heliconius melpomene	HMP500	M	ant	E-MTAB-1500
Heliconius melpomene	HMP500	M	legs	E-MTAB-1500
Heliconius melpomene	HMP500	M	mou	E-MTAB-1500
Heliconius melpomene	HMP501	M	ant	E-MTAB-1500
Heliconius melpomene	HMP501	M	legs	E-MTAB-1500
Heliconius melpomene	HMP501	M	mou	E-MTAB-1500
Heliconius melpomene	HMP502	F	ant	E-MTAB-1500
Heliconius melpomene	HMP502	F	legs	E-MTAB-1500
Heliconius melpomene	HMP502	F	mou	E-MTAB-1500
Heliconius melpomene	HMP503	F	ant	E-MTAB-1500
Heliconius melpomene	HMP503	F	legs	E-MTAB-1500
Heliconius melpomene	HMP503	F	mou	E-MTAB-1500
Heliconius melpomene	HMP514	F	ant	E-MTAB-6249
Heliconius melpomene	HMP514	F	head	E-MTAB-6249
Heliconius melpomene	HMP514	F	legs	E-MTAB-6249
Heliconius melpomene	HMP514	F	mou	E-MTAB-6249
Heliconius melpomene	HMP515	M	ant	E-MTAB-6249
Heliconius melpomene	HMP515	M	head	E-MTAB-6249
Heliconius melpomene	HMP515	M	legs	E-MTAB-6249

Heliconius melpomene	HMP515	M	mou	E-MTAB-6249

Table S3.2. List of surveyed phototransduction genes in 8 genomes

Gene	Dm	Am	Ag	Tc	Hm	Ms	Bm	Dp
Arrestin 1 (Arr1)	1	1	1	1	1	1	1	1
Arrestin 2 (Arr2)	1	1	1	1	1	1	1	1
kurtz (Krz)	1	1	1	1	1	1	1	1
Phorestin-2-like	1	1	1	1	1	1	1	1
Na/Ca-exchange protein (Calx)	1	1	1	1	1	1	1	1
CDP diglyceride synthetase (CdsA)	1	1	1	1	1	1	1	1
G protein alpha49B (Galpha49B)	1	1	1	1	1	1	1	1
G protein beta-subunit 76C (Gbeta76C)	1	1	1	1	1	1	1	1
G protein gamma30A (Ggamma30A)	1	1	1	1	1	1	1	1
G protein-coupled receptor kinase 1 (Gprk1)	1	1	1	1	1	1	1	1
G protein-coupled receptor kinase 2 (Gprk2)	1	1	1	1	1	1	1	1
inactivation no afterpotential D (InaD)	1	1	1	1	1	1	1	1
neither inactivation nor afterpotential C (ninaC)	1	1	1	1	1	1	1	1
ninaC2	0	1	0	0	1	1	1	1
ninaG	1	1	1	1	1	1	1	1
no receptor potential A (norpA)	1	1	1	1	1	1	1	1
blue opsin (BRh)	1	0	0	0	1	1	1	1
ultraviolet opsin (UVRh)	2	0	0	1	2	1	1	1
long wavelength-sensitive opsin (LWRh)	3	0	0	1	1	1	2	1
rhodopsin 7 (Rh7)	1	0	0	0	1	1	1	1
pteropsin	0	1	2	1	1	1	1	1
unclassified opsin-like	0	0	0	0	1	1	0	1
RPE-retinal G protein-coupled receptor-like (RGR-like)	0	0	0	0	1	1	1	1
Phosphatidylinositol synthase (Pis)	1	1	1	1	1	1	1	1
Phospholipase D (Pld)	1	1	1	1	1	1	1	1
retinal degeneration A (rgA)	1	1	1	1	1	1	1	1
retinal degeneration B (rdgB)	1	1	1	1	1	1	1	1
retinal degeneration B beta (rdgBbeta)	1	1	1	1	1	1	1	1
retinal degeneration C (rdgC)	1	0	1	1	1	1	1	1
transient receptor potential (trp)	1	1	1	1	1	1	1	1
trpgamma (trpg)	1	1	1	1	1	1	1	1
trp-like (trpl)	1	1	1	1	1	1	1	1
retinal dehydrogenase B (pdhb)	1	1	1	1	1	1	1	1
RabX4	1	1	1	1	1	1	1	1
Drab5 (Rab5)	1	1	1	1	1	1	1	1
cacophony (cac)	1	1	1	1	1	1	1	1
Calmodulin (Cam)	1	1	1	1	1	1	1	1
Cam-similar	0	1	0	1	1	1	1	1
Cam-like	1	1	1	1	1	1	1	1
Dopa decarboxylase (Ddc)	1	1	1	1	1	1	1	1

Ddc-like	1	1	1	1	1	3	1	1
Dual oxidase (Duox)	1	1	1	1	1	1	1	1
inactivation no afterpotential E (inaE)	1	1	1	1	0	0	0	0
Diacylglycerol Lipase-beta (DAGL-beta)	0	1	1	1	1	1	1	1
Lazaro (laza)	1	0	0	0	0	0	0	0
wunen (wun)	2	0	0	0	3	3	3	3
laza-like	0	1	1	1	3	3	3	3
Sodium/potassium/calcium exchanger (Nckx30C)	1	1	1	1	1	1	1	1
ninaA	1	1	1	1	1	1	1	1
Protein C kinase 53E (PKc53E)	1	1	1	1	1	1	1	1
inactivation no afterpotential C (inaC)	1	0	0	0	0	0	0	0
shaking B (shakB)	1	1	1	1	1	1	1	1
innexin3	1	1	1	1	1	1	1	1
innexin2-like	0	0	0	1	2	1	1	1
innexin2	1	1	1	1	1	1	1	1
innexin6	1	0	0	0	0	0	0	0
innexin5	1	0	0	0	0	0	0	0
innexin4	1	0	1	0	0	0	0	0
innexin1	1	1	1	1	1	1	1	1
innexin7 (inx7)	1	1	1	1	1	1	1	1
shaker cognate b (shakerB)	1	1	1	1	1	1	1	1
Vacuolar H+ ATPase 100kD subunit 1 (Vha100-1)	1	1	1	2	1	1	1	1
Vha100-1-like	0	0	0	0	1	1	1	1
Vha100-2	1	1	1	1	1	1	1	1

Dm Drosophila melanogaster
Am Apis mellifera
Ag Anopheles gambiae
Tc Tribolium castaneum
Hm Heliconius melpomene
Ms Manduca sexta
Dp Danaus plexippus

Table S3.3. Annotated DE contigs in head vs antennae comparison

Gene ID	Associated Gene	Gene ID	Associated Gene	Gene ID	Associated Gene
comp26215	5-HT2A	comp509368	CG42750	comp23006	Ir87a
comp9102	a10	comp32431	CG4302	comp26876	Ir8a
comp444905	acj6	comp28583	CG4306	comp8056	Ir93a
comp32563	Acon	comp31612	CG43066	comp22794	Ir93a
comp33795	Act57B	comp28598	CG43078	comp33436	jv
comp30846	Actn	comp121466	CG43163	comp23689	Kaz1-ORFB
comp13934	Acyp2	comp25788	CG43163	comp33506	kcc
comp24455	Adgf-A	comp32323	CG43370	comp33156	kek5
comp29352	al	comp32292	CG4341	comp32633	kek6
comp31806	al	comp29354	CG43707	comp28725	kn
comp29864	Aldh-III	comp30880	CG43729	comp32108	Kul
comp32612	Alk	comp30150	CG43729	comp32732	l(2)01289
comp28799	alpha-Est3	comp33189	CG4386	comp32732	l(2)01289
comp15551	alpha-Est9	comp30820	CG44085	comp27453	l(2)efl
comp25257	alphaTub85E	comp22614	CG44422	comp24179	l(2)efl
comp26800	Amt	comp31635	CG45076	comp28468	l(3)72Dp
comp31671	Ance	comp31635	CG45076	comp26135	laccase2
comp33309	Ance-3	comp32531	CG45105	comp33432	lbe
comp27820	Ank2	comp22726	CG4525	comp22379	Lcch3
comp31368	AOX1	comp17464	CG4525	comp26265	Lcp65Ac
comp17607	AOX1	comp20436	CG4525	comp8435	Lip4
comp32219	AOX1	comp32995	CG4562	comp9097	Lmpt
comp32219	AOX1	comp30081	CG4576	comp9087	lush
comp393944	AOX2	comp26603	CG4753	comp23097	Mal-A4
comp31368	AOX3	comp30410	CG4822	comp23097	Mal-A8
comp31368	AOX3	comp31620	CG4945	comp32527	Mdr49
comp162171	AOX3	comp28748	CG4972	comp33482	Meltrin
comp25833	AOX3	comp104272	CG5009	comp29724	mesh
comp32219	AOX4	comp96721	CG5009	comp29288	Mf
comp30253	AQP	comp24877	CG5023	comp29526	mgl
comp26053	ara	comp31322	CG5065	comp33046	mgl
comp30803	Arr1	comp31374	CG5142	comp31912	Mhc
comp21614	Arr2	comp29619	CG5191	comp25374	mid
comp29459	AstA	comp9688	CG5435	comp25526	Mlc1
comp121787	AstC-R2	comp25982	CG5621	comp25637	Mlp84B
comp16227	b6	comp25982	CG5621	comp26631	Mp
comp7533	B9d2	comp31437	CG5665	comp22276	Mp20
comp25682	BBS8	comp26638	CG5756	comp33698	mRpS34
comp32350	beat-IIa	comp32995	CG5789	comp22537	MSBP
comp22531	beat-IIIa	comp11275	CG5921	comp33012	msn

comp29113	beat-IIIb	comp26245	CG5964	comp29083	mspo
comp28872	beat-IV	comp32752	CG6084	comp30093	Msr-110
comp17826	beat-IV	comp23449	CG6287	comp32300	mub
comp2317	beat-VI	comp28260	CG6329	comp26218	Mur89F
comp30011	betaTub56D	comp24263	CG6414	comp33012	Myo28B1
comp20143	bol	comp25404	CG6472	comp20723	mys
comp457640	bsh	comp9977	CG6484	comp32295	nAChRalpha1
comp26704	bt	comp26113	CG6638	comp29356	nAChRalpha2
comp33771	bt	comp30531	CG6723	comp22329	nAChRalpha3
comp25672	bt	comp21663	CG6726	comp29161	nAChRbeta1
comp33771	bt	comp20732	CG6800	comp30645	nAChRbeta2
comp33694	Cad96Ca	comp32775	CG6867	comp33698	Nckx30C
comp33074	Calx	comp30189	CG6938	comp23741	ND-18
comp32772	CarT	comp30189	CG6938	comp33696	Ndae1
comp33584	cas	comp23604	CG7149	comp33696	Ndae1
comp15278	Cat	comp23164	CG7236	comp28863	Ndg
comp31653	cd	comp25097	CG7236	comp32373	ndl
comp23662	CDC45L	comp28946	CG7442	comp33036	Nep4
comp33079	Cdk5alpha	comp25732	CG7458	comp33149	NimA
comp33135	cep290	comp20562	CG7497	comp22610	ninaA
comp25804	CG10026	comp32665	CG7509	comp23799	ninaB
comp19520	CG10026	comp32642	CG7609	comp32517	ninaB
comp31868	CG10026	comp31766	CG7675	comp33012	ninaC
comp24263	CG10175	comp28742	CG7742	comp30705	ninaC
comp31796	CG10348	comp28742	CG7742	comp30705	ninaC
comp32411	CG10384	comp32650	CG7837	comp26822	ninaG
comp20770	CG10939	comp26626	CG7985	comp30836	Nlg2
comp33497	CG10960	comp115998	CG8028	comp29395	Nlg3
comp24174	CG10960	comp32373	CG8172	comp31914	Nmdar1
comp31866	CG11000	comp33466	CG8249	comp28323	Nmdar1
comp21383	CG1113	comp32258	CG8306	comp28323	Nmdar1
comp13356	CG11155	comp28249	CG8389	comp24083	Nmdar2
comp29736	CG11200	comp32671	CG8596	comp32751	nolo
comp33212	CG11200	comp19858	CG8654	comp60363	nompB
comp33212	CG11200	comp30821	CG8745	comp31587	norpA
comp23012	CG1126	comp30821	CG8745	comp25267	Npc2h
comp24847	CG11284	comp32736	CG8785	comp32367	Nplp1
comp17803	CG11294	comp28949	CG8888	comp29266	Nrg
comp29505	CG11318	comp33650	CG8909	comp31929	nrv2
comp33120	CG11321	comp28453	CG8916	comp25186	Oaz
comp33120	CG11321	comp30922	CG8918	comp31457	Oaz
comp24419	CG11380	comp25880	CG9119	comp12076	Obp19d
comp27538	CG11550	comp26136	CG9297	comp23820	Obp28a

comp28699	CG11550	comp21785	CG9304	comp26213	Obp69a
comp30460	CG11550	comp24997	CG9427	comp15036	Obp83a
comp29740	CG11638	comp30081	CG9447	comp26526	Obp84a
comp23975	CG11854	comp22434	CG9447	comp26513	obst-A
comp33050	CG11883	comp31335	CG9449	comp32175	obst-E
comp24582	CG12009	comp28392	CG9503	comp28764	oc
comp24831	CG12025	comp27408	CG9512	comp31552	olf186-F
comp31957	CG1208	comp25872	CG9518	comp23832	Oli
comp28978	CG1213	comp25178	CG9521	comp23935	Or13a
comp23117	CG12355	comp25178	CG9521	comp28366	Or13a
comp29477	CG12594	comp663	CG9522	comp14023	Or13a
comp29237	CG1271	comp25757	CG9541	comp20410	Or43a
comp32373	CG1273	comp31146	CG9631	comp24441	Or46a
comp21379	CG12811	comp24709	CG9631	comp25542	Or46a
comp313101	CG13055	comp26911	CG9701	comp27377	Or67c
comp30187	CG13192	comp33180	CG9701	comp20533	Or85b
comp41045	CG13282	comp28022	CG9784	comp27024	Orco
comp29922	CG13318	comp32828	CG9813	comp26705	ort
comp33471	CG13458	comp30046	CG9864	comp16597	Oseg2
comp33471	CG13458	comp25795	CG9981	comp22495	Oseg2
comp26556	CG13675	comp29758	CG9990	comp27590	Oseg5
comp31228	CG13875	comp25796	Che-13	comp25244	Osm6
comp32397	CG1402	comp33772	chp	comp32629	Papss
comp20724	CG14022	comp30843	clumsy	comp32515	Parg
comp25739	CG14024	comp28620	clumsy	comp32103	pdgy
comp15736	CG14141	comp27042	cn	comp28533	pdgy
comp32854	CG14304	comp30363	cnc	comp28533	pdgy
comp26224	CG14314	comp23506	Con	comp39976	Pdh
comp28008	CG14367	comp26487	Cpr100A	comp31846	Peritrophin-A
comp27484	CG14372	comp1089	Cpr49Aa	comp32877	Pex7
comp30846	CG14407	comp26011	Cpr49Ac	comp1234	Phk-3
comp31322	CG1441	comp32370	Cpr67B	comp26924	Phk-3
comp33070	CG1444	comp32579	cpx	comp27819	PK1-R
comp14383	CG14457	comp32517	cv-2	comp33264	Pkcdelta
comp32428	CG14535	comp23307	Cyp12b2	comp33264	Pkcdelta
comp24794	CG14655	comp23307	Cyp12c1	comp22888	Pph13
comp24794	CG14655	comp9090	Cyp303a1	comp15800	Prm
comp33227	CG14661	comp28778	Cyp49a1	comp33822	Prm
comp31375	CG14880	comp31622	Cyp4c3	comp32534	prom
comp3381	CG15120	comp31622	Cyp4c3	comp33691	Ptp36E
comp31122	CG15186	comp32164	Cyp4c3	comp16890	PTPMT1
comp32196	CG15221	comp32963	Cyp4d20	comp30415	Pxd
comp23968	CG15406	comp32963	Cyp4g15	comp31394	Pxd

comp32278	CG1544	comp32930	Cyp4s3	comp31694	pyd3
comp22005	CG15478	comp18551	Cyp6a2	comp32703	pyx
comp28360	CG15506	comp28330	Cyp6a2	comp180016	rad
comp17273	CG15537	comp27642	Cyp6a2	comp28672	RapGAP1
comp24446	CG15661	comp29497	Cyp6a2	comp30064	RBP
comp17268	CG15666	comp26685	Cyp6a2	comp19822	Rdl
comp29075	CG15701	comp25121	Cyp6d5	comp32599	RecQ5
comp32895	CG15765	comp27610	Cyp6w1	comp29054	rempA
comp27649	CG15890	comp18299	Ddc	comp33735	Rfabg
comp32607	CG1607	comp29258	Desat1	comp29668	Rh3
comp27190	CG16700	comp33382	Dfd	comp29377	Rh5
comp31565	CG1674	comp27260	Dh31	comp31862	RhoGAP102A
comp27773	CG16854	comp513	Dh44	comp31682	RIC-3
comp27773	CG16854	comp31623	dila	comp33346	rl
comp32771	CG17111	comp24120	DIP-alpha	comp27994	rod
comp32771	CG17111	comp28900	DIP-gamma	comp35092	RpL40
comp23818	CG17192	comp21878	DIP-gamma	comp33276	rst
comp184817	CG17211	comp27474	DIP-theta	comp20049	rtp
comp24518	CG17211	comp15039	DIP-zeta	comp32570	sals
comp24430	CG17270	comp30512	Dll	comp33489	santa-maria
comp29027	CG17292	comp27896	dnd	comp27443	Scgdelta
comp145899	CG17322	comp28850	Doc1	comp1715	Scgdelta
comp32538	CG17324	comp27256	Dop1R1	comp29883	Scp2
comp33339	CG17364	comp28586	dpr18	comp33318	ScpX
comp28517	CG17572	comp27880	dpr8	comp28895	Scr
comp27569	CG17599	comp27663	dpr8	comp27722	scro
comp28074	CG17664	comp26314	Drak	comp21655	scu
comp20976	CG17669	comp32424	dsb	comp27239	SERCA
comp27503	CG17739	comp17333	Dscam1	comp28390	shakB
comp29289	CG17834	comp27882	Dscam3	comp23621	Sherpa
comp32811	CG17999	comp23191	Dscam3	comp33712	sif
comp28427	CG18586	comp1742	Dscam3	comp402497	sim
comp28352	CG18641	comp4091	Dscam3	comp205025	sim
comp22794	CG1909	comp287315	Dscam3	comp32015	sl
comp30828	CG2070	comp18922	Dscam3	comp33302	Slc45-1
comp34179	CG2254	comp17822	Dscam3	comp30698	slif
comp32812	CG2269	comp326926	Dscam3	comp23764	Sln
comp32084	CG2540	comp31693	Dscam4	comp27365	sls
comp21567	CG2650	comp31693	Dscam4	comp29096	sls
comp34180	CG2650	comp319268	Dscam4	comp26145	sls
comp24427	CG2650	comp18318	Dscam4	comp32363	sls
comp29217	CG2663	comp129558	dyl	comp15832	smog
comp27885	CG2663	comp30185	Dyrk2	comp31440	Smyd4-2

comp28803	CG2663	comp29764	E(spl)mbeta-HLH	comp26395	SmydA-2
comp29636	CG2663	comp30021	E(spl)mgamma-HLH	comp33257	SmydA-4
comp30928	CG2681	comp33679	Eaat1	comp31543	SmydA-8
comp30841	CG2930	comp9107	EbpIII	comp31578	snk
comp33738	CG30274	comp25985	emp	comp31578	snk
comp29716	CG30427	comp28099	en	comp31150	Snmp1
comp20573	CG30441	comp8598	epsilonTry	comp14595	Snmp2
comp33490	CG30463	comp10262	Erk7	comp20343	sNPF
comp28312	CG30497	comp18160	Erk7	comp29503	sns
comp24709	CG31326	comp370264	erm	comp31453	sosie
comp29030	CG31344	comp30610	Ets98B	comp20042	Sox100B
comp26264	CG31549	comp27997	Exn	comp33181	Sp1
comp30745	CG3164	comp31159	eya	comp30122	Sp212
comp28572	CG31676	comp21455	eyg	comp29626	Spn77Ba
comp30811	CG3168	comp132327	eys	comp31070	Spn85F
comp23921	CG31760	comp69398	eys	comp26371	srp
comp33357	CG31869	comp18273	eys	comp28499	ss
comp21673	CG31954	comp24285	fa2h	comp32443	St4
comp23844	CG32091	comp24426	fas	comp19506	ste14
comp30648	CG32121	comp385505	fd59A	comp28763	stg1
comp31901	CG32225	comp25756	fd96Ca	comp2520	stj
comp27830	CG32354	comp56525	fd96Ca	comp1126	stj
comp29252	CG32354	comp36076	Fkbp14	comp28798	stops
comp32666	CG32432	comp9158	fln	comp28798	stops
comp26756	CG32447	comp31348	Fmo-1	comp15432	Strn-Mlck
comp27081	CG3259	comp26863	FoxK	comp15432	Strn-Mlck
comp31932	CG32668	comp33450	fus	comp33528	Syt1
comp24154	CG32698	comp27356	futsch	comp33714	tacc
comp29868	CG33098	comp24661	fw	comp254047	tey
comp31129	CG33543	comp1165	GABA-B-R3	comp28555	Tg
comp31459	CG34113	comp29267	Gad1	comp26751	Tis11
comp30733	CG34113	comp23703	Galphaq	comp27475	Tk
comp15356	CG34351	comp32027	Gat	comp28933	Tm2
comp32069	CG34357	comp35049	Gbeta76C	comp33464	tn
comp32069	CG34357	comp31146	gd	comp33464	tn
comp33596	CG34396	comp29963	Gdh	comp20280	to
comp32035	CG3630	comp31333	Ge-1	comp29696	Tollo
comp20910	CG3700	comp32042	Gel	comp26970	TpnC25D
comp25465	CG3769	comp25637	Gint3	comp28528	TpnC41C
comp14994	CG3790	comp24598	gl	comp15524	TpnC73F
comp27711	CG3812	comp3676	Gld	comp27901	Trc8
comp26377	CG3812	comp32239	GluClalpha	comp34345	Tret1-1

comp32282	CG3822	comp129538	GluRIA	comp374122	Trh
comp28979	CG3823	comp29702	Glycogenin	comp32711	trpl
comp30717	CG3829	comp168885	Gr43a	comp31882	Try29F
comp25745	CG40160	comp28334	Grip	comp25386	Tsp29Fa
comp31649	CG40485	comp22269	GstD11	comp28985	Tsp29Fb
comp32610	CG42249	comp29741	gukh	comp30391	Tsp47F
comp30801	CG42260	comp26568	Hasp	comp26311	Tsp47F
comp30801	CG42260	comp22681	Hasp	comp31662	Tsp97E
comp33660	CG42265	comp32566	Hdc	comp24219	Ugt86De
comp33738	CG42284	comp30418	HisCl1	comp32559	Ugt86Di
comp30702	CG42321	comp27528	Hmx	comp31135	unc80
comp30702	CG42321	comp29768	HnRNP-K	comp33686	Unc-89
comp32043	CG42326	comp33438	HPS4	comp21597	up
comp32235	CG42339	comp33273	inaD	comp25975	v
comp27087	CG4239	comp20113	Ipp	comp32454	VGlut
comp318773	CG42402	comp1859	Ir21a	comp32399	Vha100-2
comp33007	CG42492	comp71754	Ir21a	comp26976	Vmat
comp30375	CG42534	comp30400	Ir25a	comp9123	wat
comp32510	CG42613	comp10013	Ir31a	comp33411	wry
comp19727	CG42673	comp20821	Ir41a	comp25538	wupA
comp32807	CG42675	comp30063	Ir75d	comp31651	Zasp66
comp30885	CG42748	comp10821	Ir75d	comp25040	zyd

Table S3.4. Annotated DE contigs in head vs legs comparison

Gene ID	Associated Gene	Gene ID	Associated Gene	Gene ID	Associated Gene
comp122407	2mit	comp28802	CG4467	comp25526	Mlc1
comp26215	5-HT2A	comp30652	CG4467	comp25637	Mlp84B
comp33443	ab	comp27353	CG45002	comp29375	mmd
comp32312	Ac3	comp31635	CG45076	comp29375	mmd
comp24884	Ac3	comp31635	CG45076	comp26631	Mp
comp28440	AcCoAS	comp33678	CG45186	comp27007	Mp
comp28440	AcCoAS	comp30081	CG4576	comp24023	Mp20
comp32740	Ace	comp32294	CG4587	comp22276	Mp20
comp33087	Ace	comp33574	CG4615	comp33698	mRpS34
comp33795	Act57B	comp26603	CG4753	comp33012	msn
comp30846	Actn	comp31370	CG4797	comp29083	mspo
comp21713	Adf1	comp31620	CG4945	comp20683	mtt
comp31806	al	comp33321	CG4984	comp24243	Muc91C
comp29352	al	comp367738	CG5009	comp32759	mwh
comp28799	alpha-Est3	comp24877	CG5023	comp33502	mxt
comp33421	amon	comp456468	CG5038	comp33012	Myo28B1
comp26800	Amt	comp26077	CG5065	comp32463	mys
comp31671	Ance	comp27135	CG5326	comp24392	mys
comp32768	Ance	comp33555	CG5549	comp32295	nAChRalpha1
comp32768	Ance	comp33555	CG5549	comp29356	nAChRalpha2
comp29904	Ance-3	comp26506	CG5618	comp22329	nAChRalpha3
comp33513	Ank2	comp25982	CG5621	comp28917	nAChRalpha5
comp33513	Ank2	comp25982	CG5621	comp29712	nAChRalpha6
comp33513	Ank2	comp26638	CG5756	comp29161	nAChRbeta1
comp27820	Ank2	comp22213	CG6024	comp30645	nAChRbeta2
comp28953	Antp	comp33235	CG6236	comp31105	NaCP60E
comp30209	AOX1	comp33235	CG6236	comp31105	NaCP60E
comp29189	Aplip1	comp23025	CG6279	comp32170	NaPi-III
comp30253	AQP	comp31958	CG6282	comp32262	NaPi-III
comp33859	Argk	comp28260	CG6329	comp23885	nau
comp30803	Arr1	comp24263	CG6414	comp33698	Nckx30C
comp21614	Arr2	comp25371	CG6432	comp33696	Ndae1
comp20886	Asph	comp33121	CG6454	comp33696	Ndae1
comp20886	Asph	comp9622	CG6495	comp28863	Ndg
comp29459	AstA	comp30531	CG6723	comp31810	Neurochondrin
comp31078	Atpalpha	comp21663	CG6726	comp22610	ninaA
comp16227	b6	comp27246	CG6726	comp23799	ninaB
comp22531	beat-IIIa	comp28469	CG6765	comp30965	ninaB
comp27800	beat-IIIc	comp32775	CG6867	comp30705	ninaC
comp31896	beat-IIIc	comp23604	CG7149	comp30705	ninaC

comp19200	beat-VI	comp31241	CG7433	comp33012	ninaC
comp26157	beat-VI	comp28946	CG7442	comp26822	ninaG
comp2317	beat-VI	comp29964	CG7458	comp27289	NLaz
comp31778	beat-VI	comp32665	CG7509	comp30836	Nlg2
comp23162	Bgb	comp28692	CG7708	comp29395	Nlg3
comp30037	B-H1	comp26626	CG7985	comp26254	Nlg3
comp31377	Bili	comp30967	CG8086	comp24222	Nlg3
comp457640	bsh	comp25011	CG8192	comp31334	Nlg3
comp33373	bsk	comp32258	CG8306	comp31002	Nlg3
comp25672	bt	comp28249	CG8389	comp28323	Nmdar1
comp33771	bt	comp10890	CG8407	comp28323	Nmdar1
comp26704	bt	comp32705	CG8500	comp24083	Nmdar2
comp33771	bt	comp33381	CG8547	comp28136	noc
comp26941	bves	comp19858	CG8654	comp21563	Non2
comp32762	Bx	comp30821	CG8745	comp31587	norpA
comp27635	cac	comp30821	CG8745	comp11292	Nos
comp33694	Cad96Ca	comp31058	CG8785	comp25267	Npc2h
comp33064	CadN	comp32736	CG8785	comp32367	Nplp1
comp33064	CadN	comp28949	CG8888	comp29266	Nrg
comp33744	CadN	comp33650	CG8909	comp31929	nrv2
comp33427	Cadps	comp28453	CG8916	comp32524	nrv2
comp33427	Cadps	comp30922	CG8918	comp32524	nrv2
comp33427	Cadps	comp32008	CG9062	comp30615	nrv3
comp33427	Cadps	comp25499	CG9098	comp32342	nSyb
comp30660	CAH2	comp27556	CG9171	comp29506	nSyb
comp33074	Calx	comp26136	CG9297	comp30569	nwk
comp25632	Cam	comp17421	CG9314	comp31457	Oaz
comp33566	Cam	comp27764	CG9391	comp25186	Oaz
comp10352	Camta	comp28405	CG9436	comp12076	Obp19d
comp22829	Camta	comp30081	CG9447	comp23820	Obp28a
comp32584	Cand1	comp31335	CG9449	comp15036	Obp83a
comp27945	capt	comp28392	CG9503	comp32175	obst-E
comp32772	CarT	comp27408	CG9512	comp28764	oc
comp31653	cd	comp30535	CG9518	comp26420	Octbeta1R
comp33194	Cdep	comp29541	CG9518	comp18903	Octbeta3R
comp33079	Cdk5alpha	comp19465	CG9626	comp17728	Oct-TyrR
comp32461	CdsA	comp31146	CG9631	comp31552	olf186-F
comp31868	CG10026	comp22010	CG9636	comp23832	Oli
comp19520	CG10026	comp29482	CG9646	comp26644	onecut
comp31414	CG10089	comp31248	CG9657	comp26705	ort
comp29425	CG10137	comp26911	CG9701	comp32507	pad
comp24263	CG10175	comp31504	CG9701	comp33553	Pal2

comp23465	CG10188	comp31504	CG9701	comp32515	Parg
comp29470	CG10188	comp28022	CG9784	comp31080	Pask
comp2222	CG10345	comp33030	CG9812	comp32722	Pax
comp32411	CG10384	comp32828	CG9813	comp30635	Pde6
comp24030	CG10407	comp31087	CG9896	comp28533	pdgy
comp32264	CG10710	comp24526	CG9919	comp28533	pdgy
comp29763	CG10737	comp15437	CG9953	comp25541	Pdh
comp31752	CG10738	comp33270	chas	comp32877	Pex7
comp31752	CG10738	comp23627	ChAT	comp30576	PGAP5
comp31752	CG10738	comp32833	cher	comp24632	PGRP-LC
comp321560	CG10804	comp33772	chp	comp26924	Phk-3
comp12098	CG10804	comp32678	clumsy	comp1234	Phk-3
comp26111	CG10864	comp30886	Cngl	comp24162	Phk-3
comp29803	CG1090	comp33203	comt	comp30523	PICK1
comp20770	CG10939	comp23506	Con	comp31053	Pif1A
comp24174	CG10960	comp5482	Con	comp9171	Pis
comp28807	CG10960	comp21325	Con	comp27819	PK1-R
comp33497	CG10960	comp15046	Cow	comp15628	Pka-C1
comp31866	CG11000	comp32603	Cp110	comp33264	Pkcdelta
comp24060	CG11034	comp26487	Cpr100A	comp33264	Pkcdelta
comp13356	CG11155	comp23327	Cpr64Ab	comp30751	pod1
comp17803	CG11294	comp21375	Cpr92A	comp26206	Poxm
comp27353	CG11319	comp32579	cpx	comp22888	Pph13
comp33120	CG11321	comp33285	Cubn	comp21731	prc
comp33120	CG11321	comp23307	Cyp12b2	comp33822	Prm
comp30964	CG1136	comp23307	Cyp12c1	comp15800	Prm
comp24419	CG11380	comp9090	Cyp303a1	comp32534	prom
comp30460	CG11550	comp21913	Cyp304a1	comp135045	PsGEF
comp27538	CG11550	comp30413	Cyp4c3	comp29178	PsGEF
comp29740	CG11638	comp31622	Cyp4c3	comp16890	PTPMT1
comp9124	CG11843	comp31622	Cyp4c3	comp30415	Pxd
comp23975	CG11854	comp32930	Cyp4s3	comp31694	pyd3
comp24582	CG12009	comp29497	Cyp6a2	comp32703	pyx
comp28978	CG1213	comp27642	Cyp6a2	comp33518	qvr
comp31746	CG1213	comp903	D	comp24200	Rab26
comp32088	CG12173	comp32770	Dab	comp33355	Rab3
comp29477	CG12594	comp32770	Dab	comp33677	Rab3-GEF
comp30379	CG12796	comp32901	Dap160	comp33677	Rab3-GEF
comp21499	CG12814	comp18299	Ddc	comp33677	Rab3-GEF
comp30113	CG12858	comp26510	Desat1	comp30306	rad
comp465916	CG12950	comp29258	Desat1	comp28672	RapGAP1
comp29637	CG12950	comp33382	Dfd	comp29488	ras

comp26790	CG12950	comp27260	Dh31	comp30064	RBP
comp31878	CG1299	comp513	Dh44	comp32249	rdgB
comp24035	CG13003	comp33371	didum	comp19822	Rdl
comp32662	CG13248	comp24120	DIP-alpha	comp30888	Rdl
comp41045	CG13282	comp20667	DIP-delta	comp30888	Rdl
comp29922	CG13318	comp32358	DIP-eta	comp27308	ref(2)P
comp6830	CG13375	comp7075	DIP-eta	comp28293	Reph
comp31920	CG13397	comp21878	DIP-gamma	comp33735	Rfabg
comp31361	CG13409	comp28900	DIP-gamma	comp30181	rg
comp31361	CG13409	comp293181	DIP-theta	comp33311	rg
comp33471	CG13458	comp27474	DIP-theta	comp27739	Rgk1
comp33471	CG13458	comp19528	disco-r	comp29668	Rh3
comp29891	CG13634	comp33053	Doa	comp29377	Rh5
comp31211	CG13676	comp27256	Dop1R1	comp33422	rhea
comp31228	CG13875	comp1592	Dop1R2	comp25906	rho-5
comp15736	CG14141	comp30959	Dop2R	comp28412	RhoGAP100F
comp27068	CG14207	comp26466	DopEcR	comp31862	RhoGAP102A
comp22499	CG14312	comp27057	dpr1	comp31682	RIC-3
comp26224	CG14314	comp29605	dpr12	comp33639	Rim
comp24399	CG14321	comp28213	dpr12	comp33639	Rim
comp27484	CG14372	comp6997	dpr13	comp24476	r-l
comp21931	CG14372	comp28586	dpr18	comp29138	robo3
comp30846	CG14407	comp24366	dpr20	comp19401	robo3
comp33043	CG14439	comp1695	dpr4	comp32791	Rop
comp32957	CG14446	comp18415	dpr4	comp35092	RpL40
comp14383	CG14457	comp627	dpr7	comp31469	RpS27A
comp26605	CG14507	comp27663	dpr8	comp20049	rtp
comp32428	CG14535	comp27880	dpr8	comp33769	RyR
comp22134	CG1461	comp29678	dpr9	comp32570	sals
comp24794	CG14655	comp886	dpy	comp31888	Sap47
comp24794	CG14655	comp27685	Dr	comp23168	sca
comp33227	CG14661	comp19816	drd	comp23168	sca
comp2855	CG14669	comp18808	Drep2	comp19711	scaf
comp33591	CG14762	comp22889	Drep2	comp29545	Sclp
comp31375	CG14880	comp31942	Drip	comp29883	Scp2
comp16327	CG14964	comp30465	Drl-2	comp33318	ScpX
comp23676	CG14984	comp26794	Dscam2	comp28895	Scr
comp24894	CG1504	comp27882	Dscam3	comp27722	scro
comp32196	CG15221	comp20483	Dscam3	comp31963	Sema-1a
comp26995	CG15221	comp623858	Dscam3	comp32127	Sema-2a
comp29779	CG15412	comp4091	Dscam3	comp26593	Sep4
comp22005	CG15478	comp287315	Dscam3	comp27239	SERCA

comp28360	CG15506	comp326926	Dscam3	comp33167	sev
comp17273	CG15537	comp1742	Dscam3	comp22210	sff
comp30342	CG1561	comp20876	Dscam3	comp30110	sff
comp26151	CG15630	comp17822	Dscam3	comp33653	Shab
comp29746	CG15760	comp18922	Dscam3	comp28390	shakB
comp32895	CG15765	comp20556	Dscam3	comp29515	Shaw
comp33403	CG15894	comp23191	Dscam3	comp4044	Shawl
comp32607	CG1607	comp26794	Dscam3	comp30598	shi
comp32145	CG16711	comp319268	Dscam4	comp25331	shot
comp31565	CG1674	comp31693	Dscam4	comp29410	sif
comp25289	CG16786	comp31693	Dscam4	comp33712	sif
comp28537	CG16798	comp30185	Dyrk2	comp17662	SIFaR
comp21844	CG16799	comp32230	dysc	comp32159	Sik2
comp16245	CG16817	comp33231	e	comp205025	sim
comp20177	CG1695	comp20574	ea	comp402497	sim
comp23818	CG17192	comp29711	ea	comp32015	sl
comp24518	CG17211	comp24674	eag	comp28452	sl
comp184817	CG17211	comp24674	eag	comp28452	sl
comp31004	CG17221	comp9107	EbpIII	comp16327	S-Lap1
comp24430	CG17270	comp31580	Ef1alpha100E	comp33302	Slc45-1
comp27262	CG17271	comp28938	Eip63F-1	comp31733	Sln
comp145899	CG17322	comp31535	Eip74EF	comp26145	sls
comp33339	CG17364	comp25749	Eip78C	comp29096	sls
comp23395	CG17364	comp24597	Elk	comp32363	sls
comp28936	CG17386	comp26553	Ent3	comp27365	sls
comp30247	CG17646	comp370264	erm	comp24649	smog
comp28074	CG17664	comp31674	ETHR	comp15832	smog
comp20976	CG17669	comp31538	Ets65A	comp31543	SmydA-8
comp9082	CG17684	comp27997	Exn	comp30467	Snap25
comp26723	CG17739	comp24285	fa2h	comp31578	snk
comp27503	CG17739	comp24426	fas	comp31578	snk
comp32452	CG17739	comp23115	FASN1	comp15239	sNPF-R
comp32452	CG17739	comp26446	fat-spondin	comp29503	sns
comp32593	CG17928	comp385505	fd59A	comp31453	sosie
comp32584	CG18304	comp56525	fd96Ca	comp28194	SoxN
comp32584	CG18304	comp25756	fd96Ca	comp33181	Sp1
comp28352	CG18641	comp27693	Fhos	comp24796	Sp1
comp35248	CG18815	comp12672	Fife	comp30122	Sp212
comp22794	CG1909	comp29517	Fife	comp23748	Sp212
comp28328	CG2016	comp33525	Fit1	comp31059	SP2637
comp29984	CG2064	comp9158	fln	comp78712	Spn77Ba
comp30828	CG2070	comp31348	Fmo-1	comp28587	sqa

comp34179	CG2254	comp24528	fne	comp28587	sqa
comp30302	CG2258	comp22384	fne	comp23065	sr
comp32812	CG2269	comp31844	Fnta	comp27631	ssh
comp28614	CG2556	comp31844	Fnta	comp27631	ssh
comp21567	CG2650	comp32598	form3	comp32395	st
comp34180	CG2650	comp33239	foxo	comp32443	St4
comp29217	CG2663	comp32894	Frq2	comp19506	ste14
comp30169	CG2663	comp32591	ftz-f1	comp1126	stj
comp25126	CG2767	comp30758	FucTA	comp2520	stj
comp33331	CG30069	comp22846	Fur2	comp33598	stj
comp28791	CG30069	comp31351	Fur2	comp40780	stnA
comp31498	CG30116	comp33450	fus	comp40780	stnB
comp29716	CG30427	comp32434	futsch	comp28798	stops
comp30204	CG31030	comp31770	fwe	comp28798	stops
comp31771	CG3108	comp19450	GABA-B-R1	comp15432	Strn-Mlck
comp23312	CG31191	comp29502	GABA-B-R2	comp15432	Strn-Mlck
comp27062	CG31221	comp32236	GABA-B-R2	comp30819	Syn
comp28572	CG31676	comp1165	GABA-B-R3	comp29944	Synj
comp23921	CG31760	comp320333	GABA-B-R3	comp33528	Syt1
comp33357	CG31869	comp29267	Gad1	comp33670	Syt7
comp22453	CG32052	comp23703	Galphaq	comp25842	Syt7
comp33541	CG32052	comp32027	Gat	comp30633	t
comp30648	CG32121	comp35049	Gbeta76C	comp33714	tacc
comp21285	CG32204	comp27247	gce	comp32392	tadr
comp31901	CG32225	comp31146	gd	comp32392	tadr
comp33429	CG32354	comp27237	Gfrl	comp31624	Task7
comp32666	CG32432	comp31534	GIIIspla2	comp31579	TBC1D16
comp33536	CG32485	comp25637	Gint3	comp280713	Tdc2
comp24776	CG32544	comp24598	gl	comp23893	Tdc2
comp24154	CG32698	comp30079	Gld	comp28265	Teh1
comp32882	CG32815	comp27107	glob1	comp33652	Ten-m
comp31256	CG3285	comp24561	GluRIA	comp254047	tey
comp29868	CG33098	comp31235	gprs	comp28555	Tg
comp27267	CG33203	comp168885	Gr43a	comp26751	Tis11
comp31129	CG33543	comp31928	grn	comp27475	Tk
comp30169	CG33966	comp27326	Gs1l	comp31591	Tl
comp33600	CG3409	comp17612	Gsc	comp29886	Tm1
comp30733	CG34113	comp15284	GstO3	comp28933	Tm2
comp32028	CG34113	comp15391	GstO3	comp33464	tn
comp15356	CG34351	comp29649	GstO3	comp33464	tn
comp31892	CG34354	comp32593	Gyc76C	comp26266	tok
comp32069	CG34357	comp26568	Hasp	comp32315	Tom40

comp32069	CG34357	comp22681	Hasp	comp25274	Tom40
comp33258	CG34370	comp30059	Hayan	comp32441	tou
comp32270	CG34384	comp32431	Hexo1	comp31278	tow
comp33596	CG34396	comp32580	hgo	comp28528	TpnC41C
comp33751	CG34417	comp30418	HisCl1	comp21557	TpnC41C
comp33751	CG34417	comp6123	hiw	comp15524	TpnC73F
comp22422	CG34417	comp33200	Hk	comp32808	trbd
comp28981	CG3609	comp27528	Hmx	comp28327	trbl
comp28981	CG3609	comp29091	how	comp27901	Trc8
comp32035	CG3630	comp33438	HPS4	comp24960	Treh
comp17188	CG3655	comp29005	Hr3	comp27788	Tret1-2
comp31096	CG3662	comp32973	Hsp68	comp27788	Tret1-2
comp20910	CG3700	comp31378	IA-2	comp22011	Tret1-2
comp27965	CG3703	comp27980	ics	comp24925	Tret1-2
comp14994	CG3790	comp28907	igl	comp32711	trpl
comp28995	CG3795	comp26237	ImpL2	comp22582	Trx-2
comp26377	CG3812	comp33273	inaD	comp28985	Tsp29Fb
comp32282	CG3822	comp33569	Indy	comp25027	Tsp47F
comp32678	CG3822	comp26492	InR	comp31662	Tsp97E
comp28979	CG3823	comp26492	InR	comp190370	tup
comp31379	CG3860	comp20113	Ipp	comp22760	Tusp
comp33573	CG3940	comp33304	Irk1	comp17305	Tusp
comp25745	CG40160	comp33304	Irk1	comp17967	Tusp
comp31054	CG40470	comp19926	Itgbn	comp30954	tutl
comp31649	CG40485	comp19165	jim	comp14988	Uev1A
comp30942	CG4213	comp27330	jim	comp31324	Ugt86Dd
comp30801	CG42260	comp20456	jp	comp24219	Ugt86De
comp30801	CG42260	comp27242	Jra	comp33755	unc-104
comp20733	CG42261	comp33506	kcc	comp30611	unc-13
comp261998	CG42261	comp22611	KCNQ	comp25433	unc-5
comp33660	CG42265	comp31010	kek3	comp30173	unc-5
comp32043	CG42326	comp28433	kek6	comp90252	unc79
comp122569	CG42329	comp28725	kn	comp33020	unc79
comp32235	CG42339	comp28410	krz	comp26191	unc79
comp32749	CG42346	comp32108	Kul	comp30594	unc79
comp32749	CG42346	comp32732	l(2)01289	comp31135	unc80
comp29172	CG42361	comp32732	l(2)01289	comp24542	unc80
comp30555	CG42390	comp27453	l(2)efl	comp30658	unc80
comp28927	CG42404	comp32603	l(2)efl	comp29514	Unc-89
comp30670	CG42450	comp24179	l(2)efl	comp33686	Unc-89
comp33007	CG42492	comp18440	l(2)k09913	comp21597	up
comp16000	CG42594	comp27141	l(3)neo38	comp16447	Usp30

comp32510	CG42613	comp26135	laccase2	comp29638	uzip
comp32807	CG42675	comp33743	Lar	comp28226	VAChT
comp30885	CG42748	comp33432	lbe	comp28105	VGAT
comp509368	CG42750	comp22379	Lcch3	comp32454	VGlut
comp30560	CG42750	comp8435	Lip4	comp26976	Vmat
comp32081	CG42795	comp33959	Lmpt	comp32949	Vps13
comp28583	CG4306	comp32240	LPCAT	comp32791	Vps52
comp31612	CG43066	comp32240	LPCAT	comp32791	Vps52
comp28598	CG43078	comp32994	LpR1	comp30448	vvl
comp25788	CG43163	comp31987	Lrrk	comp33215	w
comp30083	CG43366	comp30633	lush	comp25100	wal
comp32292	CG4341	comp20736	mAChR-B	comp33463	wit
comp30144	CG43693	comp33482	Meltrin	comp119162	Wnt5
comp29354	CG43707	comp416798	mesh	comp33411	wry
comp252687	CG43707	comp31808	MESK2	comp25538	wupA
comp30150	CG43729	comp27247	Met	comp33534	yellow-f
comp25836	CG43737	comp13705	mfr	comp19833	yip2
comp31941	CG43778	comp29526	mgl	comp33080	Zasp52
comp29086	CG43897	comp31912	Mhc	comp31651	Zasp66
comp33502	CG44085	comp25374	mid	comp20571	Zasp67
comp29974	CG44247	comp31563	Mitf	comp28920	zyd

Table S3.5. Annotated DE contigs in head vs mouth parts comparison.

Gene ID	Associated Gene	Gene ID	Associated Gene	Gene ID	Associated Gene
comp122407	2mit	comp104272	CG5009	comp23097	Mal-A4
comp22620	5-HT1B	comp26077	CG5065	comp23097	Mal-A8
comp33443	ab	comp3603	CG5065	comp26124	Mctp
comp32312	Ac3	comp26220	CG5162	comp33482	Meltrin
comp24884	Ac3	comp27135	CG5326	comp25974	Membrin
comp28440	AcCoAS	comp27762	CG5390	comp31808	MESK2
comp28440	AcCoAS	comp9084	CG5390	comp29526	mgl
comp32740	Ace	comp23842	CG5390	comp23997	miple1
comp27682	Acer	comp33555	CG5549	comp29307	MKP-4
comp25481	Acp65Aa	comp33555	CG5549	comp28209	Moe
comp21713	Adf1	comp26506	CG5618	comp27036	mol
comp11590	AIMP3	comp26638	CG5756	comp26631	Mp
comp31806	al	comp32849	CG5789	comp27007	Mp
comp29352	al	comp33353	CG5883	comp32597	Mp
comp29864	Aldh-III	comp22213	CG6024	comp24023	Mp20
comp15551	alpha-Est9	comp33549	CG6052	comp33698	mRpS34
comp33421	amon	comp26945	CG6106	comp22537	MSBP
comp29904	Ance-3	comp31832	CG6142	comp33012	msn
comp33513	Ank2	comp17543	CG6154	comp29083	mspo
comp33513	Ank2	comp31998	CG6271	comp20683	mtt
comp33513	Ank2	comp23025	CG6279	comp32300	mub
comp27820	Ank2	comp31000	CG6294	comp22556	Mur89F
comp33515	Antp	comp28260	CG6329	comp32759	mwh
comp32219	AOX1	comp29219	CG6414	comp33723	Myo10A
comp32219	AOX1	comp21111	CG6414	comp32776	Myo10A
comp32219	AOX4	comp30531	CG6723	comp33012	Myo28B1
comp20923	ap	comp21663	CG6726	comp32463	mys
comp29189	Aplip1	comp27246	CG6726	comp27360	Naam
comp30263	aret	comp28469	CG6765	comp32295	nAChRalpha1
comp30803	Arr1	comp32775	CG6867	comp29356	nAChRalpha2
comp21614	Arr2	comp23604	CG7149	comp22329	nAChRalpha3
comp29459	AstA	comp29238	CG7231	comp28837	nAChRalpha4
comp32955	att-ORFA	comp28946	CG7442	comp28917	nAChRalpha5
comp32350	beat-IIa	comp29964	CG7458	comp29712	nAChRalpha6
comp22531	beat-IIIa	comp24758	CG7461	comp29161	nAChRbeta1
comp27800	beat-IIIc	comp20562	CG7497	comp30645	nAChRbeta2
comp26157	beat-VI	comp32665	CG7509	comp31105	NaCP60E
comp23162	Bgb	comp32849	CG7627	comp31105	NaCP60E
comp30037	B-H1	comp31766	CG7675	comp29408	nahoda
comp31377	Bili	comp30905	CG7702	comp32170	NaPi-III

comp276518	bnl	comp28692	CG7708	comp32262	NaPi-III
comp29930	bond	comp15709	CG7777	comp33698	Nckx30C
comp457640	bsh	comp26626	CG7985	comp33696	Ndae1
comp33373	bsk	comp30967	CG8086	comp33696	Ndae1
comp33771	bt	comp32258	CG8306	comp28863	Ndg
comp25672	bt	comp28249	CG8389	comp33036	Nep4
comp33771	bt	comp32380	CG8398	comp33149	NimA
comp32762	Bx	comp29011	CG8420	comp22610	ninaA
comp22647	C901	comp29011	CG8420	comp30705	ninaC
comp33638	Ca-beta	comp9780	CG8483	comp30705	ninaC
comp31219	cac	comp33381	CG8547	comp33012	ninaC
comp31219	cac	comp27096	CG8646	comp26822	ninaG
comp27635	cac	comp28740	CG8785	comp30054	NLaz
comp33694	Cad96Ca	comp32736	CG8785	comp30836	Nlg2
comp33064	CadN	comp28949	CG8888	comp29395	Nlg3
comp33064	CadN	comp33650	CG8909	comp26254	Nlg3
comp33744	CadN	comp28453	CG8916	comp31334	Nlg3
comp33427	Cadps	comp32008	CG9062	comp28323	Nmdar1
comp33427	Cadps	comp25499	CG9098	comp28323	Nmdar1
comp33427	Cadps	comp27556	CG9171	comp31914	Nmdar1
comp33427	Cadps	comp30500	CG9220	comp24083	Nmdar2
comp33074	Calx	comp27764	CG9391	comp21563	Non2
comp25632	Cam	comp28405	CG9436	comp31587	norpA
comp10352	Camta	comp31335	CG9449	comp25888	Nos
comp22829	Camta	comp27408	CG9512	comp33895	Npc2b
comp30274	caps	comp31308	CG9572	comp33895	Npc2b
comp27945	capt	comp31847	CG9628	comp25267	Npc2h
comp32772	CarT	comp24709	CG9631	comp32367	Nplp1
comp33584	cas	comp22010	CG9636	comp31929	nrv2
comp31653	cd	comp18152	CG9701	comp32524	nrv2
comp33194	Cdep	comp29523	CG9701	comp32524	nrv2
comp33079	Cdk5alpha	comp28022	CG9784	comp30615	nrv3
comp32461	CdsA	comp33030	CG9812	comp33391	Nrx-1
comp30016	CG10019	comp32828	CG9813	comp32342	nSyb
comp31868	CG10026	comp28745	CG9917	comp29506	nSyb
comp31414	CG10089	comp24526	CG9919	comp30569	nwk
comp29425	CG10137	comp15437	CG9953	comp31457	Oaz
comp23465	CG10188	comp29758	CG9990	comp25186	Oaz
comp29470	CG10188	comp33270	chas	comp25217	Obp19d
comp33531	CG10211	comp23627	ChAT	comp23820	Obp28a
comp28998	CG10268	comp32833	cher	comp28764	oc
comp2222	CG10345	comp33772	chp	comp26420	Octbeta1R

comp32502	CG10362	comp28036	Cib2	comp18903	Octbeta3R
comp24030	CG10407	comp30886	Cngl	comp17728	Oct-TyrR
comp31224	CG10638	comp33203	comt	comp31552	olf186-F
comp30183	CG10663	comp5482	Con	comp23832	Oli
comp12098	CG10804	comp21325	Con	comp26644	onecut
comp321560	CG10804	comp15046	Cow	comp26705	ort
comp29803	CG1090	comp28281	Cpr11B	comp32507	pad
comp20770	CG10939	comp102666	Cpr11B	comp33553	Pal2
comp24174	CG10960	comp30129	Cpr49Aa	comp32515	Parg
comp28807	CG10960	comp1089	Cpr49Aa	comp31080	Pask
comp32096	CG10960	comp37756	Cpr49Aa	comp30635	Pde6
comp33497	CG10960	comp27321	Cpr49Ae	comp32103	pdgy
comp31866	CG11000	comp32370	Cpr67B	comp32360	pdgy
comp13356	CG11155	comp20061	Cpr73D	comp32464	pdgy
comp29736	CG11200	comp21375	Cpr92A	comp32877	Pex7
comp17803	CG11294	comp32579	cpx	comp30576	PGAP5
comp27353	CG11319	comp29592	Cralbp	comp25233	PGRP-LB
comp29740	CG11638	comp31679	CtsB1	comp1234	Phk-3
comp9124	CG11843	comp33484	cwo	comp30523	PICK1
comp23975	CG11854	comp33484	cwo	comp33061	pio
comp24582	CG12009	comp9090	Cyp303a1	comp16973	pip
comp24831	CG12025	comp6427	Cyp4d1	comp9171	Pis
comp6321	CG12071	comp32963	Cyp4d20	comp27819	PK1-R
comp31746	CG1213	comp32963	Cyp4g15	comp24037	ple
comp28230	CG1213	comp32930	Cyp4s3	comp26206	Poxm
comp32088	CG12173	comp33146	Cyp6g2	comp22888	Pph13
comp23117	CG12355	comp31083	Cyp9f2	comp21731	prc
comp30113	CG12858	comp35808	Cyt-b5	comp32534	prom
comp29637	CG12950	comp903	D	comp29178	PsGEF
comp31878	CG1299	comp30246	dac	comp33691	Ptp36E
comp29029	CG13024	comp32901	Dap160	comp16890	PTPMT1
comp313101	CG13055	comp30335	Dat	comp26798	PTPMT1
comp30187	CG13192	comp18299	Ddc	comp29331	Pu
comp32662	CG13248	comp27559	Desat1	comp24123	pug
comp31363	CG13293	comp33382	Dfd	comp30346	Pvf3
comp29922	CG13318	comp513	Dh44	comp31694	pyd3
comp6830	CG13375	comp24120	DIP-alpha	comp32703	pyx
comp31361	CG13409	comp248607	DIP-beta	comp33518	qvr
comp31361	CG13409	comp20667	DIP-delta	comp24200	Rab26
comp29891	CG13634	comp32358	DIP-eta	comp33355	Rab3
comp19176	CG1368	comp28900	DIP-gamma	comp33677	Rab3-GEF
comp31228	CG13875	comp293181	DIP-theta	comp33677	Rab3-GEF

comp25739	CG14024	comp15039	DIP-zeta	comp33677	Rab3-GEF
comp15736	CG14141	comp23518	dlg1	comp30306	rad
comp22499	CG14312	comp30512	Dll	comp28672	RapGAP1
comp26224	CG14314	comp28850	Doc1	comp30064	RBP
comp24399	CG14321	comp27256	Dop1R1	comp19822	Rdl
comp27484	CG14372	comp1592	Dop1R2	comp30888	Rdl
comp33043	CG14439	comp30959	Dop2R	comp30888	Rdl
comp33070	CG1444	comp26466	DopEcR	comp31865	Rep
comp26605	CG14507	comp27057	dpr1	comp33735	Rfabg
comp32428	CG14535	comp29605	dpr12	comp27739	Rgk1
comp2855	CG14669	comp6997	dpr13	comp29668	Rh3
comp33591	CG14762	comp28586	dpr18	comp29377	Rh5
comp31375	CG14880	comp24366	dpr20	comp28412	RhoGAP100F
comp24894	CG1504	comp1695	dpr4	comp33639	Rim
comp22005	CG15478	comp18415	dpr4	comp33639	Rim
comp27645	CG15531	comp27663	dpr8	comp24390	rn
comp17273	CG15537	comp27880	dpr8	comp29138	robo3
comp21810	CG15628	comp29678	dpr9	comp32791	Rop
comp28148	CG15651	comp25099	dpy	comp20049	rtp
comp9149	CG15739	comp22025	dpy	comp32570	sals
comp29746	CG15760	comp19173	dpy	comp33489	santa-maria
comp32895	CG15765	comp886	dpy	comp31888	Sap47
comp32494	CG15890	comp21177	dpy	comp23168	sca
comp27649	CG15890	comp21177	dpy	comp23168	sca
comp877281	CG15890	comp27922	dpy	comp30188	scaf
comp33403	CG15894	comp27685	Dr	comp1715	Scgdelta
comp32607	CG1607	comp18808	Drep2	comp29545	Sclp
comp28537	CG16798	comp22889	Drep2	comp29883	Scp2
comp21844	CG16799	comp30465	Drl-2	comp27722	scro
comp16245	CG16817	comp26794	Dscam2	comp25835	sei
comp27773	CG16854	comp1130	Dscam3	comp31963	Sema-1a
comp27773	CG16854	comp326926	Dscam3	comp28165	Sep2
comp463180	CG17211	comp4091	Dscam3	comp26593	Sep4
comp24518	CG17211	comp1742	Dscam3	comp30854	sev
comp184817	CG17211	comp18922	Dscam3	comp33167	sev
comp24430	CG17270	comp20483	Dscam3	comp30110	sff
comp27262	CG17271	comp27882	Dscam3	comp22210	sff
comp29027	CG17292	comp20876	Dscam3	comp33653	Shab
comp13745	CG17292	comp26794	Dscam3	comp29515	Shaw
comp23395	CG17364	comp23191	Dscam3	comp30598	shi
comp33339	CG17364	comp319268	Dscam4	comp33770	shot
comp28517	CG17572	comp31693	Dscam4	comp33712	sif

comp20976	CG17669	comp31693	Dscam4	comp29410	sif
comp26723	CG17739	comp30115	Duox	comp17662	SIFaR
comp32452	CG17739	comp30185	Dyrk2	comp402497	sim
comp32452	CG17739	comp32230	dysc	comp205025	sim
comp33579	CG17999	comp29764	E(spl)mbeta-HLH	comp32571	SK
comp28092	CG18301	comp29711	ea	comp32015	sl
comp28352	CG18641	comp20574	ea	comp28452	sl
comp35248	CG18815	comp21506	ea	comp28452	sl
comp22794	CG1909	comp33679	Eaat1	comp33302	Slc45-1
comp28328	CG2016	comp29522	Eaat2	comp31733	Sln
comp28257	CG2150	comp29522	Eaat2	comp29096	sls
comp30302	CG2258	comp24674	eag	comp24649	smog
comp32812	CG2269	comp24674	eag	comp15832	smog
comp32597	Cg25C	comp9107	EbpIII	comp31543	SmydA-8
comp33254	Cg25C	comp25031	EbpIII	comp31578	snk
comp34180	CG2650	comp33577	ec	comp31578	snk
comp7853	CG2650	comp31580	Ef1alpha100E	comp14595	Snmp2
comp28803	CG2663	comp25985	emp	comp15239	sNPF-R
comp29217	CG2663	comp20928	en	comp29503	sns
comp30169	CG2663	comp28099	en	comp25654	Snup
comp25126	CG2767	comp26553	Ent3	comp17963	Sos
comp30204	CG31030	comp370264	erm	comp31453	sosie
comp25705	CG31097	comp31674	ETHR	comp29790	Sox102F
comp27062	CG31221	comp31538	Ets65A	comp22181	Sox21b
comp24709	CG31326	comp27997	Exn	comp28194	SoxN
comp25113	CG31326	comp29055	Faa	comp33181	Sp1
comp33586	CG31522	comp24426	fas	comp23567	Sp212
comp29547	CG31523	comp32460	Fas2	comp30122	Sp212
comp28572	CG31676	comp26446	fat-spondin	comp23748	Sp212
comp30811	CG3168	comp385505	fd59A	comp27148	Sp7
comp14787	CG31728	comp25756	fd96Ca	comp21430	Spat
comp23921	CG31760	comp33525	Fit1	comp31207	Spn100A
comp22453	CG32052	comp36076	Fkbp14	comp32602	Spn42Da
comp29902	CG32052	comp24528	fne	comp20550	Spn42Da
comp33541	CG32052	comp22384	fne	comp28532	Spn42Dc
comp31418	CG32206	comp33239	foxo	comp19729	Spn77Ba
comp31901	CG32225	comp26738	Fpps	comp29626	Spn77Ba
comp33429	CG32354	comp32894	Frq2	comp28961	Spn88Ea
comp27830	CG32354	comp29806	fu12	comp26371	srp
comp32666	CG32432	comp30758	FucTA	comp30486	ssp6
comp33536	CG32485	comp22846	Fur2	comp30364	stck
comp24776	CG32544	comp31351	Fur2	comp19506	ste14

comp24154	CG32698	comp33450	fus	comp28763	stg1
comp32882	CG32815	comp27356	futsch	comp2520	stj
comp20798	CG3285	comp19450	GABA-B-R1	comp1126	stj
comp29868	CG33098	comp3181	GABA-B-R1	comp33598	stj
comp32577	CG33143	comp29502	GABA-B-R2	comp40780	stnA
comp27267	CG33203	comp32236	GABA-B-R2	comp40780	stnB
comp9126	CG33229	comp1165	GABA-B-R3	comp28798	stops
comp31129	CG33543	comp320333	GABA-B-R3	comp28798	stops
comp30169	CG33966	comp29267	Gad1	comp32933	svp
comp33277	CG33970	comp23703	Galphaq	comp32119	svr
comp33600	CG3409	comp32027	Gat	comp33672	Swim
comp30733	CG34113	comp30905	GATAd	comp30819	Syn
comp31892	CG34354	comp15353	GatB	comp30582	Syngr
comp32069	CG34357	comp35049	Gbeta76C	comp33528	Syt1
comp32069	CG34357	comp32042	Gel	comp33670	Syt7
comp32270	CG34384	comp24598	gl	comp32392	tadr
comp33751	CG34417	comp30079	Gld	comp32392	tadr
comp33751	CG34417	comp32239	GluClalpha	comp31579	TBC1D16
comp391035	CG34461	comp24561	GluRIA	comp280713	Tdc2
comp26059	CG3618	comp21639	Glut1	comp28265	Teh1
comp20910	CG3700	comp32298	GNBP3	comp33652	Ten-m
comp27965	CG3703	comp31235	gprs	comp24642	Tep3
comp14994	CG3790	comp168885	Gr43a	comp254047	tey
comp26377	CG3812	comp31928	grn	comp28555	Tg
comp27711	CG3812	comp29728	grnd	comp27475	Tk
comp32282	CG3822	comp22365	Gs2	comp31591	Tl
comp31379	CG3860	comp29649	GstO3	comp33464	tn
comp19297	CG40006	comp26568	Hasp	comp33464	tn
comp25745	CG40160	comp22681	Hasp	comp28464	Toll-6
comp32499	CG4133	comp32566	Hdc	comp28464	Tollo
comp22640	CG4168	comp31891	Hexo1	comp31278	tow
comp23516	CG4168	comp30418	HisCl1	comp15524	TpnC73F
comp32610	CG42249	comp6123	hiw	comp24960	Treh
comp30801	CG42260	comp33200	Hk	comp27788	Tret1-2
comp30801	CG42260	comp26795	Hmgcr	comp27788	Tret1-2
comp261998	CG42261	comp28228	Hmgs	comp248258	Tret1-2
comp20733	CG42261	comp27528	Hmx	comp31559	Trp1
comp32180	CG42269	comp29091	how	comp32711	trpl
comp27686	CG42269	comp29005	Hr3	comp31882	Try29F
comp30702	CG42321	comp30742	Hs3st-B	comp28638	Try29F
comp30702	CG42321	comp31688	hth	comp25463	Tsp39D
comp32043	CG42326	comp31378	IA-2	comp25027	Tsp47F

comp32235	CG42339	comp31950	ICA69	comp31662	Tsp97E
comp29172	CG42361	comp33760	if	comp17305	Tusp
comp28927	CG42404	comp28907	igl	comp17967	Tusp
comp30670	CG42450	comp27226	ImpL2	comp33755	unc-104
comp33007	CG42492	comp33273	inaD	comp30611	unc-13
comp16000	CG42594	comp20113	Ipp	comp33726	unc-13
comp31994	CG42594	comp19926	Itgbn	comp22367	unc-4
comp32510	CG42613	comp19555	Itgbn	comp25433	unc-5
comp32807	CG42675	comp27330	jim	comp30173	unc-5
comp30372	CG42732	comp27242	Jra	comp33020	unc79
comp30885	CG42748	comp23689	Kaz1-ORFB	comp26191	unc79
comp30560	CG42750	comp33506	kcc	comp30594	unc79
comp32081	CG42795	comp30969	kek1	comp31135	unc80
comp28583	CG4306	comp31010	kek3	comp24542	unc80
comp31612	CG43066	comp29187	ken	comp30658	unc80
comp355137	CG4313	comp28296	Khc-73	comp33686	Unc-89
comp25788	CG43163	comp28725	kn	comp29514	Unc-89
comp29354	CG43707	comp26896	knrl	comp23432	unpg
comp252687	CG43707	comp32108	Kul	comp21597	up
comp30150	CG43729	comp33953	l(2)34Fc	comp25975	v
comp25836	CG43737	comp27453	l(2)efl	comp28226	VAChT
comp31941	CG43778	comp24179	l(2)efl	comp32349	veil
comp29974	CG44247	comp28397	l(3)mbn	comp33296	veil
comp22614	CG44422	comp33743	Lar	comp28105	VGAT
comp30652	CG4467	comp33432	lbe	comp32454	VGlut
comp28802	CG4467	comp22379	Lcch3	comp26976	Vmat
comp27353	CG45002	comp26265	Lcp65Ac	comp32949	Vps13
comp31635	CG45076	comp26230	Lcp65Ac	comp32791	Vps52
comp31635	CG45076	comp26867	Lcp65Ad	comp32791	Vps52
comp32294	CG4587	comp27235	Lcp65Ad	comp9123	wat
comp8079	CG4598	comp25161	Lcp65Ad	comp29384	wat
comp28932	CG4678	comp32197	Lim1	comp24093	wat
comp28932	CG4678	comp33959	Lmpt	comp33463	wit
comp32119	CG4678	comp32107	loh	comp119162	Wnt5
comp26603	CG4753	comp32240	LPCAT	comp33411	wry
comp33579	CG4830	comp32240	LPCAT	comp25538	wupA
comp32274	CG4928	comp9087	lush	comp33080	Zasp52
comp27481	CG4998	comp20736	mAChR-B	comp31651	Zasp66
comp96721	CG5009	comp14220	Mal-A1	comp28920	zyd

Table S3.6. Functional enrichment of differentially expressed genes.

Cluster	Term	Count	PVal	Genes	FDR
Annotation Cluster A	detection of light stimulus	15	4.07E-09	CALX, SANTA-MARIA, STOPS, GBETA76C, NINAC, NINAB, INAD, ARR2, ARR1, NINAA, TRPL, NORPA, RH3, RH5, SHAKB	5.45E-06
Enrichment Score: 6.115	detection of abiotic stimulus	15	1.54E-08	CALX, SANTA-MARIA, STOPS, GBETA76C, NINAC, NINAB, INAD, ARR2, ARR1, NINAA, TRPL, NORPA, RH3, RH5, SHAKB	2.06E-05
	detection of external stimulus	15	5.01E-08	CALX, SANTA-MARIA, STOPS, GBETA76C, NINAC, NINAB, INAD, ARR2, ARR1, NINAA, TRPL, NORPA, RH3, RH5, SHAKB	6.70E-05
Annotation Cluster B	Glucose-methanol-choline oxidoreductase	7	1.36E-05	CG9521, GLD, CG9518, CG9522, CG9512, NINAG, CG9503	0.008
Enrichment Score: 4.748	Alcohol_oxidase	7	1.57E-05	CG9521, GLD, CG9518, CG9522, CG9512, NINAG, CG9503	0.002
	Glucose-methanol-choline oxidoreductase, N-terminal	7	2.18E-05	CG9521, GLD, CG9518, CG9522, CG9512, NINAG, CG9503	0.012
Annotation Cluster C	regulation of rhodopsin mediated signaling pathway	7	7.09E-06	NINAC, INAD, ARR2, ARR1, STOPS, NORPA, GBETA76C	0.009
Enrichment Score: 4.699	deactivation of rhodopsin mediated signaling	7	7.09E-06	NINAC, INAD, ARR2, ARR1, STOPS, NORPA, GBETA76C	0.009
	rhodopsin mediated phototransduction	7	1.77E-05	NINAC, INAD, ARR2, ARR1, STOPS, NORPA, GBETA76C	0.023
Annotation Cluster D	detection of light stimulus involved in visual perception	8	8.44E-06	NINAC, INAD, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	0.011

Enrichment Score: 4.183	detection of light stimulus involved in sensory perception	8	1.20E-05	NINAC, INAD, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	0.016
	detection of stimulus involved in sensory perception	8	0.002786	NINAC, INAD, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	0.976
Annotation Cluster E	Homeobox	15	1.82E-05	ARA, CG11294, BSH, DFD, PPH13, DLL, SCRO, LBE, EN, HMX, ACJ6, SCR, AL, OC, EYG	0.004
Enrichment Score: 4.107	Homeobox, conserved site	15	6.31E-05	ARA, CG11294, BSH, DFD, PPH13, DLL, SCRO, LBE, EN, HMX, ACJ6, SCR, AL, OC, EYG	0.036
	Homeobox	15	1.10E-04	ARA, CG11294, BSH, DFD, PPH13, DLL, SCRO, LBE, EN, HMX, ACJ6, SCR, AL, OC, EYG	0.062
Annotation Cluster F	regulation of rhodopsin mediated signaling pathway	10	2.63E-09	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, RDGB, NORPA, GBETA76C	4.19E-06
Enrichment Score: 7.841	deactivation of rhodopsin mediated signaling	10	2.63E-09	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, RDGB, NORPA, GBETA76C	4.19E-06
	rhodopsin mediated phototransduction	10	1.37E-08	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, RDGB, NORPA, GBETA76C	2.19E-05
Annotation Cluster G	detection of light stimulus involved in visual perception	11	1.71E-08	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, TRPL, RDGB, NORPA, GBETA76C	2.73E-05
Enrichment Score: 6.385	detection of light stimulus involved in sensory perception	11	3.00E-08	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, TRPL, RDGB, NORPA, GBETA76C	4.78E-05
	detection of stimulus involved	11	1.37E	NINAC, CAMTA, INAD, CAM, ARR2, ARR1,	0.196

	in sensory perception		-04	STOPS, TRPL, RDGB, NORPA, GBETA76C	
Annotation Cluster H	skeletal muscle tissue development	10	3.65E -05	WUPA, WIT, FUTSCH, DAP160, FRQ2, MLP84B, SLS, UNC-104, GAD1, LBE	0.057
Enrichment Score: 4.080	striated muscle tissue development	10	1.13E -04	WUPA, WIT, FUTSCH, DAP160, FRQ2, MLP84B, SLS, UNC-104, GAD1, LBE	0.166
	muscle tissue development	10	1.39E -04	WUPA, WIT, FUTSCH, DAP160, FRQ2, MLP84B, SLS, UNC-104, GAD1, LBE	0.199
Annotation Cluster I	Homeobox	17	1.70E -05	CG11294, B-H1, BSH, GSC, DFD, ONECUT, PPH13, SCRO, TUP, LBE, HMX, DR, SCR, AL, VVL, OC, ANTP	0.005
Enrichment Score: 3.974	Homeobox, conserved site	17	5.59E -05	CG11294, B-H1, BSH, GSC, DFD, ONECUT, PPH13, SCRO, TUP, LBE, HMX, DR, SCR, AL, VVL, OC, ANTP	0.041
	Homeobox	17	1.04E -04	CG11294, B-H1, BSH, GSC, DFD, ONECUT, PPH13, SCRO, TUP, LBE, HMX, DR, SCR, AL, VVL, OC, ANTP	0.075
Annotation Cluster J	deactivation of rhodopsin mediated signaling	9	4.86E -08	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, NORPA, GBETA76C	8.03E -05
Enrichment Score: 6.693	regulation of rhodopsin mediated signaling pathway	9	4.86E -08	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, NORPA, GBETA76C	8.03E -05
	rhodopsin mediated phototransductio n	9	1.92E -07	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, NORPA, GBETA76C	3.17E -04
Annotation Cluster K	Homeobox	20	4.99E -08	CG11294, B-H1, BSH, ONECUT, DFD, PPH13, SCRO, DLL, UNC-4, LIM1, UNPG, LBE, EN, HMX, HTH, DR, AL, OC, AP, ANTP	1.28E -05
Enrichment Score: 6.283	Homeobox, conserved site	20	3.30E -07	CG11294, B-H1, BSH, ONECUT, DFD, PPH13, SCRO, DLL, UNC-4, LIM1, UNPG, LBE, EN, HMX, HTH, DR, AL, OC, AP, ANTP	2.34E -04

	Homeobox	20	7.46E-07	CG11294, B-H1, BSH, ONECUT, DFD, PPH13, SCRO, DLL, UNC-4, LIM1, UNPG, LBE, EN, HMX, HTH, DR, AL, OC, AP, ANTP	5.30E-04
Annotation Cluster L	detection of light stimulus involved in visual perception	10	1.69E-07	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	2.79E-04
Enrichment Score: 5.567	detection of light stimulus involved in sensory perception	10	2.74E-07	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	4.52E-04
	detection of stimulus involved in sensory perception	10	4.32E-04	NINAC, CAMTA, INAD, CAM, ARR2, ARR1, STOPS, TRPL, NORPA, GBETA76C	0.510
Annotation Cluster M	skeletal muscle tissue development	10	2.26E-05	FAS2, WUPA, WIT, FUTSCH, DAP160, FRQ2, SLS, UNC-104, GAD1, LBE	0.037
Enrichment Score: 4.283	striated muscle tissue development	10	7.13E-05	FAS2, WUPA, WIT, FUTSCH, DAP160, FRQ2, SLS, UNC-104, GAD1, LBE	0.111
	muscle tissue development	10	8.76E-05	FAS2, WUPA, WIT, FUTSCH, DAP160, FRQ2, SLS, UNC-104, GAD1, LBE	0.135
Annotation Cluster N	ion channel activity	24	2.76E-05	OLF186-F, SEI, PYX, CG8916, LCCH3, NMDAR1, CG11155, CAC, SK, CG4587, SHAB, NACP60E, ORT, GLUCLALPHA, NMDAR2, SHAW, RDL, CA-BETA, HISCL1, CNGL, TRPL, CG42260, EAG, CG3822	0.015
Enrichment Score: 4.281	substrate specific channel activity	24	4.34E-05	OLF186-F, SEI, PYX, CG8916, LCCH3, NMDAR1, CG11155, CAC, SK, CG4587, SHAB, NACP60E, ORT, GLUCLALPHA, NMDAR2, SHAW, RDL, CA-BETA, HISCL1, CNGL, TRPL, CG42260, EAG, CG3822	0.024
	channel activity	24	7.92E-05	OLF186-F, SEI, PYX, CG8916, LCCH3, NMDAR1, CG11155, CAC, SK, CG4587, SHAB, NACP60E, ORT, GLUCLALPHA, NMDAR2, SHAW, RDL, CA-BETA, HISCL1, CNGL, TRPL, CG42260, EAG, CG3822	0.043

DE head vs. antennea
DE head vs. legs
DE head vs. mouth

Table S3.7. Annotated contigs commonly upregulated in heads.

Gene ID	Associated Gene	Gene ID	Associated Gene	Gene ID	Associated Gene
comp30803	Arr1	comp32665	CG7509	comp33698	Nckx30C
comp21614	Arr2	comp26626	CG7985	comp33696	Ndae1
comp29459	AstA	comp28249	CG8389	comp33696	Ndae1
comp22531	beat-IIIa	comp33650	CG8909	comp28863	Ndg
comp457640	bsh	comp28453	CG8916	comp22610	ninaA
comp33074	Calx	comp31335	CG9449	comp30705	ninaC
comp32772	CarT	comp27408	CG9512	comp30705	ninaC
comp31653	cd	comp28022	CG9784	comp33012	ninaC
comp33079	Cdk5alpha	comp32828	CG9813	comp26822	ninaG
comp24174	CG10960	comp33772	chp	comp30836	Nlg2
comp31866	CG11000	comp32579	cpx	comp29395	Nlg3
comp13356	CG11155	comp513	Dh44	comp28323	Nmdar1
comp17803	CG11294	comp24120	DIP-alpha	comp28323	Nmdar1
comp29740	CG11638	comp28900	DIP-gamma	comp24083	Nmdar2
comp24582	CG12009	comp27256	Dop1R1	comp31587	norpA
comp15736	CG14141	comp28586	dpr18	comp25267	Npc2h
comp26224	CG14314	comp27663	dpr8	comp32367	Nplp1
comp27484	CG14372	comp27880	dpr8	comp31929	nrv2
comp32428	CG14535	comp1742	Dscam3	comp28764	oc
comp22005	CG15478	comp18922	Dscam3	comp23832	Oli
comp17273	CG15537	comp23191	Dscam3	comp26705	ort
comp32895	CG15765	comp27882	Dscam3	comp32515	Parg
comp32607	CG1607	comp326926	Dscam3	comp32877	Pex7
comp184817	CG17211	comp4091	Dscam3	comp22888	Pph13
comp24518	CG17211	comp31693	Dscam4	comp32534	prom
comp24430	CG17270	comp31693	Dscam4	comp28672	RapGAP1
comp33339	CG17364	comp319268	Dscam4	comp30064	RBP
comp20976	CG17669	comp30185	Dyrk2	comp19822	Rdl
comp22794	CG1909	comp370264	erm	comp33735	Rfabg
comp32812	CG2269	comp24426	fas	comp29668	Rh3
comp23921	CG31760	comp385505	fd59A	comp29377	Rh5
comp31901	CG32225	comp1165	GABA-B-R3	comp20049	rtp
comp32666	CG32432	comp29267	Gad1	comp29883	Scp2
comp24154	CG32698	comp23703	Galphaq	comp27722	scro
comp31129	CG33543	comp32027	Gat	comp33712	sif
comp30733	CG34113	comp35049	Gbeta76C	comp205025	sim
comp32069	CG34357	comp24598	gl	comp402497	sim
comp32069	CG34357	comp168885	Gr43a	comp32015	sl
comp14994	CG3790	comp22681	Hasp	comp33302	Slc45-1

comp26377	CG3812	comp26568	Hasp	comp15832	smog
comp32282	CG3822	comp30418	HisCl1	comp31543	SmydA-8
comp30801	CG42260	comp27528	Hmx	comp29503	sns
comp30801	CG42260	comp33273	inaD	comp19506	ste14
comp32235	CG42339	comp20113	Ipp	comp1126	stj
comp33007	CG42492	comp33506	kcc	comp2520	stj
comp32510	CG42613	comp28725	kn	comp28798	stops
comp30885	CG42748	comp32108	Kul	comp28798	stops
comp28583	CG4306	comp27453	l(2)efl	comp33528	Syt1
comp31612	CG43066	comp22379	Lcch3	comp254047	tey
comp29354	CG43707	comp9109	LWRh	comp28555	Tg
comp30150	CG43729	comp26631	Mp	comp27475	Tk
comp26603	CG4753	comp33698	mRpS34	comp32711	trpl
comp26638	CG5756	comp33012	msn	comp31662	Tsp97E
comp28260	CG6329	comp33012	Myo28B1	comp31135	unc80
comp30531	CG6723	comp32295	nAChRalpha1	comp32454	VGlut
comp21663	CG6726	comp29356	nAChRalpha2	comp26976	Vmat
comp32775	CG6867	comp22329	nAChRalpha3	comp33411	wry
comp23604	CG7149	comp29161	nAChRbeta1		
comp28946	CG7442	comp30645	nAChRbeta2		

CONCLUSIONS

Although some Lepidoptera (moths and butterflies) have been shown to possess color vision, not much is known about the genes involved in light processing in this insect group. The focus of previous studies is the opsin gene family, which encode proteins that initiate the phototransduction cascade (Arikawa et al. 2003; Kelber et al. 2003; Xu et al. 2013; McCulloch et al. 2017). In Lepidoptera, opsin genes are evolving by duplications, losses and changes in amino acid sequence(Briscoe & Chittka 2001; Briscoe 2001; Frentiu et al. 2007; Briscoe et al. 2010; Feuda et al. 2016). In insects, the phototransduction cascade is only well-characterized in *Drosophila* (Hardie 2001; Montell 2012; Hardie & Juusola 2015). Although a lot of work is done investigating opsin evolution in Lepidoptera, we do not yet know much about the underlying mechanisms by which these opsins transduce light information to the brain. Additional candidate genes involved in lepidopteran vision can be identified and studied by annotating their sequences with vision-related functions using an organism that is well-studied such as *Drosophila* (Speiser et al. 2014). A limitation in using genome to genome searches is that the identification of homologs does not tell us if the genes are functional nor when and where they are expressed. A solution to this constraint is to search transcriptomes for the presence of genes and, if possible, to evaluate different tissues for levels of expression. Absolute expression will show whether the gene is functional and expressed in the tissues being surveyed. We expect vision-related genes to be expressed in retinal tissue. For my dissertation, I benefited from RNA-Sequencing data to explore the evolution and expression of vision-related genes in moths and butterflies.

I began my dissertation by exploring the role of plasticity on vision-related genes. Using head RNA-Seq data, I compared gene expression between sexes and seasonal forms of a phenotypically plastic butterfly, *Bicyclus anynana*. *B. anynana* eyes vary in eye size between sexes and seasonal forms, and opsin expression varies between seasonal forms but only in females (Everett et al. 2012). In my study, I found more genes differentially expressed between seasonal forms than between sexes. Differentially expressed genes between seasonal forms had functions related to insect cuticle, retinaldehyde binding, chitin metabolism, odorant and juvenile hormone binding, sugar transport, and dimerization activity and binding. Differentially expressed genes not enriched for function but related to vision were genes involved in eye development and pigmentation. These results provide a list of eye development genes potentially contributing to differences in eye size between seasonal forms. To verify the influence of these genes, their expression should be compared between phenotypes at earlier points in development. Moreover, differentially expressed eye development and eye pigmentation genes, and genes annotated with visual function, showed a larger effect of season on female visual transcription. I conclude that season causes plasticity in head gene expression altering vision-related genes annotated with eye development and pigmentation functions, with a larger effect on female transcription (Macias-Muñoz et al. 2016).

In chapter 1, I annotated the *B. anynana* transcriptome with *Drosophila* gene symbols and gene ontology terms. This method allowed for the identification of homologs of genes that carry out eye development, eye pigmentation, and phototransduction in *Drosophila*. Many of these genes were not differentially expressed between heads using a transcriptome-wide approach. I visualized candidate vision gene relative expression

between sexes and seasonal forms, but this represented only the expression in adult head tissue. A comparison between tissue types might support a role in vision if the genes are upregulated in heads relative to other tissue types. Nonetheless, one functional enrichment cluster that could function in vision was the retinaldehyde binding cluster. Genes in this cluster have a CRAL-TRIO domain, which is a conserved N-terminal structural domain of proteins that transport hydrophobic molecules (Panagabko et al. 2003). Two members of the CRAL-TRIO domain gene family have a role in vision by transporting the chromophore (Wang & Montell 2005; Wu et al. 2006). While there is an expansion of this gene family in Lepidoptera, we do not know if any member of this gene family transports the chromophore molecule in Lepidoptera.

For my next chapter, I tested whether any member of the rapidly-evolving CRAL-TRIO domain gene family had a function in butterfly vision. I used the butterfly *Heliconus melpomene* to characterize the molecular evolution of CRAL-TRIO domain genes and their expression patterns across four tissue types. First, I found that many genes were located in tandem with potential copy number variation. I detected a *Heliconius*-specific expansion with the largest number of copy number variation at this location in the genome. However, these genes were not found in our transcriptome nor expressed in the tissues that we sampled thus we cannot predict whether they are transcribed and what their function might be. Next, by doing transcription-wide differential expression analysis and plotting the expression for the CRAL-TRIO domain genes, I identified one gene highly expressed and upregulated in *H. melopomene* heads relative to antennae, legs, and mouth parts. The head upregulated gene *Hme CTD31* was present as a single copy. Immunohistochemistry showed that the protein encoded by *Hme CTD31* is found in primary pigment, secondary pigment,

and tracheal cells. In addition, sequencing protein that fluoresce in a native gel, potentially binding a pigment molecule, suggested that Hme CTD31 is a good match to the proteins found in the fluorescent band. The results suggest that a non-orthologous gene to *Drosophila pinta* takes on the role of chromophore transport in butterflies (Macias-Muñoz et al. 2017).

For my final chapter, I again used phylogenetics and transcriptomics to explore the molecular evolution and expression of genes potentially involved in the phototransduction cascade of moths and butterflies. I sought to investigate which genes were potentially conserved between *Drosophila* and Lepidopteran, which were Lepidoptera-specific, and which varied between moths and butterflies. First, I identified and annotated genes that were upregulated in *H. melpomene* heads relative to legs, antennae, and mouth parts in 3 pairwise comparisons. These genes were functionally enriched for light detection and regulation of rhodopsin signaling. Visual inspection of the genes annotated with *Drosophila* gene symbols revealed that many *Drosophila* phototransduction homologs were upregulated in *H. melpomene* heads. This suggested potential functional gene conservation. To verify, I used sequences of experimentally tested phototransduction genes in *Drosphila melanogaster* to search for homologs in insect genomes of *Anopheles gambiae*, *Apis mellifera*, *Tribolium casteum*, *Bombyx mori*, *Manduca sexta*, *Danaus plexippus* and *Heliconius melpomene*. I also searched *de novo* transcriptomes of *M. sexta* and *H. melpomene* to curate genome annotations. In instances where I found Lepidopteran-specific duplications, I used RNA-Seq libraries from *M. sexta* heads and *H. melpomene* heads, antennae, legs and mouth parts to quantify the expression of paralogs in these tissues. Our results suggest that there are no consistent differences between diurnal butterflies and nocturnal moths in

phototransduction genes. In addition, most genes are conserved between Lepidoptera and *Drosophila* except in a few cases where a Lepidopteran-specific duplication is a candidate vision-related genes.

Overall this dissertation aims to explore the role of plasticity and gene duplication on butterfly visual system evolution. I focus on gene expression as a proxy for potential function in eye physiology. My findings in Chapter 1 suggest that expression of vision-related genes in a phenotypically plastic butterfly is plastic for eye development and eye pigmentation genes, with a larger effect on females. A limitation of this project is that we sampled adults. To verify which genes have a role in driving divergent eye phenotypes between seasons, gene expression should be quantified at different time points. In Chapters 2 and 3, I used phylogenetics and transcriptomics to survey gene duplications in insect phototransduction genes and conserved visual function between *Drosophila* and Lepidoptera. This study is the first to investigate the phototransduction cascade in Lepidoptera. Future studies on lepidopteran phototransduction should focus on turning off candidate genes and testing for loss of signal transduction. Further, exploring these genes across species and habitats might provide insights into how visual systems are evolving in response to changing environments.

REFERENCES

Arikawa K, Mizuno S, Kinoshita M, Stavenga DG. 2003. Coexpression of two visual pigments in a photoreceptor causes an abnormally broad spectral sensitivity in the eye of the butterfly *Papilio xuthus.* J. Neurosci. 23:4527–4532.

Briscoe a D. 2001. Functional diversification of lepidopteran opsins following gene duplication. Mol. Biol. Evol. 18:2270–2279.

Briscoe AD et al. 2010. Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. Proc. Natl. Acad. Sci. U. S. A. 107:3628–33. doi: 10.1073/pnas.0910085107.

Briscoe AD, Chittka L. 2001. The evolution of color vision in insects. Annu. Rev. Entomol. 46:471–510.

Everett A, Tong X, Briscoe AD, Monteiro A. 2012. Phenotypic plasticity in opsin expression in a butterfly compound eye complements sex role reversal. BMC Evol. Biol. 12:232. doi: 10.1186/1471-2148-12-232.

Feuda R, Marle F, Bentley MA, Holland PWH. 2016. Conservation, duplication, and divergence of five opsin genes in insect evolution. Genome Biol. Evol. 8:579–587. doi: 10.5287/bod-leian.

Frentiu FD, Bernard GD, Sison-Mangus MP, Van Zandt Brower A, Briscoe AD. 2007. Gene duplication is an evolutionary mechanism for expanding spectral diversity in the long-wavelength photopigments of butterflies. Mol. Biol. Evol. 24:2016–2028. doi: 10.1093/molbev/msm132.

Hardie RC. 2001. Phototransduction in *Drosophila melanogaster*. J. Exp. Biol. 204:3403–3409.

Hardie RC, Juusola M. 2015. Phototransduction in *Drosophila*. Curr. Opin. Neurobiol. 34:37–45. doi: 10.1016/j.conb.2015.01.008.

Kelber A, Balkenius A, Warrant EJ. 2003. Colour vision in diurnal and nocturnal hawkmoths. Integr. Comp. Biol. 43:571–579.

Macias-Muñoz A, McCulloch KJ, Briscoe AD. 2017. Copy number variation and expression analysis reveals a non-orthologous *pinta* gene family member involved in butterfly vision. Genome Biol. Evol. 9:3398–3412. doi: 10.1093/gbe/evx230.

Macias-Muñoz A, Smith G, Monteiro A, Briscoe AD. 2016. Transcriptome-wide differential gene expression in *Bicyclus anynana* butterflies: Female vision-related genes are more plastic. Mol. Biol. Evol. 33:79–92. doi: 10.1093/molbev/msv197.

McCulloch KJ et al. 2017. Sexual dimorphism and retinal mosaic diversification following the evolution of a violet receptor in butterflies. Mol. Biol. Evol. 34:2271–2284. doi: 10.1093/molbev/msx163.

Montell C. 2012. *Drosophila* visual transduction. Trends Neurosci. 35:356–363. doi: 10.1016/j.tins.2012.03.004.Drosophila.

Panagabko C et al. 2003. Ligand specificity in the CRAL-TRIO protein family. Biochemistry. 42:6467–6474. doi: 10.1021/bi034086v.

Speiser DI et al. 2014. Using phylogenetically-informed annotation (PIA) to search for light-interacting genes in transcriptomes from non-model organisms. BMC Bioinformatics. 15:350. doi: 10.1186/s12859-014-0350-x.

Wang T, Montell C. 2005. Rhodopsin formation in *Drosophila* is dependent on the PINTA retinoid-binding protein. J. Neurosci. 25:5187–94.

Wu Z et al. 2006. CRALBP ligand and protein interactions. Adv. Exp. Med. Biol. 572:477–483.

Xu P et al. 2013. The evolution and expression of the moth visual opsin family. PLoS One. 8:e78140. doi: 10.1371/journal.pone.0078140.