2-4-2019

The Developmental Evolution of the Novel Treehopper Helmet

Cera Fisher

University of Connecticut - Storrs, cera.fisher@uconn.edu

Follow this and additional works at: https://opencommons.uconn.edu/dissertations

Recommended Citation

https://opencommons.uconn.edu/dissertations/2063
Understanding morphological diversity requires understanding the developmental basis for the origin of novelty. In my dissertation, I use the treehopper helmet to investigate the origin of novelty. This 3-dimensional outgrowth of the prothoracic dorsal body wall has been molded by natural selection into myriad elaborate forms and differentiates treehoppers from their closest relatives. The helmet is hypothesized to have originated by modulation of the ancestral body wall patterning network, or alternatively by co-option of the wing or leg patterning network. I tested these hypotheses using a comparative approach in three species from the order Hemiptera: Entylia carinata, an emerging model treehopper that bears the novel helmet; Homalodisca vitripennis, a related leafhopper (Cicadellidae) that retains the plesiomorphic condition; and Oncopeltus fasciatus (Lygaeidae), a tractable lab organism that is evolutionarily equidistant to the former species.

In Chapter 1, I analyze tissue-specific transcriptomes (RNA-seq) of Entylia and Homalodisca. In clustering analyses of differentially expressed genes, the treehopper pronotum (helmet) is most similar to their wings, while the leafhopper pronotum is most similar to its serial homologue the mesonotum. These results support the wing patterning co-option hypothesis for the origin of the treehopper helmet. In Chapter 2, I use gene coexpression networks to identify coregulated sets of genes that are associated with the helmet and other tissues. These results further support a relationship between the helmet and genes with known roles in wing development. This chapter also presents a method for robustly distinguishing tissue-specific signals from sample effects. In Chapter 3, I
characterize the roles of seven transcription factor encoding genes in metamorphic
development of *Oncopeltus* using RNA interference (RNAi). These genes (*apterous, araucan/caupolican, homothorax, nubbin, tailup, tiptop, vestigial*) are all associated with helmet or wing clusters. Their RNAi phenotypes support the existence of a versatile developmental module for flat body wall outgrowths in insects that may predate wings. In Chapter 4, I present the first report of successful RNAi in a treehopper, *Entylia*, paving the way for further fruitful investigation of this fascinating insect’s novel morphology.
The Developmental Evolution of the Novel Treehopper Helmet
Cera Ruth-Marie Fisher

B.S., Arizona State University, 2009
M.S., Arizona State University, 2011

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2019
Copyright by

Cera Ruth-Marie Fisher
 APPROVAL PAGE

Doctor of Philosophy Dissertation

The Developmental Evolution of the Novel Treehopper Helmet

Presented by

Cera Ruth-Marie Fisher, B.S., M.S.

Major Advisor

__________________________________________________________________________
Elizabeth L. Jockusch

Associate Advisor

__________________________________________________________________________
Charles Henry

Associate Advisor

__________________________________________________________________________
David Wagner

Associate Advisor

__________________________________________________________________________
Jill Wegrzyn

Associate Advisor

__________________________________________________________________________
Yaowu Yuan

University of Connecticut
2019
Acknowledgements

Many hands contributed to make this dissertation possible. It would be too difficult to name them all. I wish to particularly recognize my domestic partner in crime, Thomas Fisher, for being my friend for almost twenty years. My advisor, Elizabeth L. Jockusch, has provided me with a phenomenal space in which to learn, and to grow from being a science enthusiast to being truly a scientist. Thank you to David Wagner for teaching me about insects. Thank you to Jill Wegrzyn for being so enthusiastic about this project and for never letting me feel too hopeless. Thank you to the many wonderful graduate students of EEB, without whom I would never have finished. Thank you to Hannah Ralicki, Holly Brown, and Sarah Elizabeth McLauren for being the best best-friends anyone could ask for. Thank you to the other wonderful graduate students who made up the GEU-UAW organizing committee; you made me proud. Thank you to my mother, Mary Ruth Womble Lawrence, R.N., I.B.C.L.C, for inspiring me to add more letters to my name.

Thank you, last but not least, to my three amazing little sisters—Cecily, Chelsea, and Chaeney—the other pieces of me living their best lives out in the world.
Dedication

This dissertation is dedicated to the memory of Chuck Wayne Lawrence, M.D., b. October 24, 1951, d. October 24, 2016.

*M.D. means “my daddy”, or something just the same.*
## TABLE OF CONTENTS

Approval page........................................................................................................................... iii

Acknowledgements................................................................................................................... iv

Dedication.................................................................................................................................. v

Chapter 1

The treehopper helmet, a morphological novelty, evolved via co-option of wing-patterning genes......................................................................................................................... 1

   Literature cited..................................................................................................................... 27

   Tables................................................................................................................................. 41

   Figures............................................................................................................................... 45

Chapter 2

Understanding the evolution of novelty as changes in regulatory networks via weighted gene co-expression analysis.......................................................................................... 54

   Literature cited..................................................................................................................... 74

   Tables................................................................................................................................. 80

   Figures............................................................................................................................... 90

Chapter 3

The evolution of adult form via an ancestral, evolutionarily versatile body wall margin module in *Oncopeltus fasciatus*............................................................................................................. 112

   Literature cited..................................................................................................................... 134

   Tables................................................................................................................................. 141

   Figures............................................................................................................................... 145
Chapter 4

Development of functional genomic analysis tools in an emerging model treehopper, *Entylia carinata*.....................................................................................................................153

Literature cited...........................................................................................................171

Tables.........................................................................................................................182

Figures........................................................................................................................185

Appendix A

Hierarchical clustering of *E. carinata* RNA-seq data without forewing samples.....194

Appendix B

TPM normalization between species.........................................................................196

Appendix C

Linear discriminant analysis with MLSeq.................................................................203
Chapter 1: The treehopper helmet, a morphological novelty, evolved via cooption of wing-patterning genes

Abstract

Understanding the origin of novelty is crucial to explaining the diversity of life. In arthropods, the body wall has served as a repeated source of novelty: legs, gills, beetle horns, and possibly even wings originated as novel body wall outgrowths. In the hemipteran family Membracidae (treehoppers), an ancestrally flat part of the dorsal body wall of the wingless first thoracic segment (the pronotum) was transformed into a three-dimensional cuticular sculpture (the helmet). Treehopper helmets have subsequently been molded by natural selection into a diverse array of shapes aiding predator defense. Developmentally, the helmet is hypothesized to have originated by elaboration of the ancestral body wall patterning network, or alternatively by co-option of the wing or leg patterning network. These three hypotheses make specific predictions about patterns of similarity in gene expression across tissue types and species. We tested these predictions using comparative transcriptomics of eight tissues in nymphal Entylia carinata (a treehopper) and Homalodisca vitripennis (a leafhopper). In the leafhopper, the pronotum and mesonotum are most similar, as would be predicted of serial homologues. In treehoppers, however, gene expression in the developing helmet is most similar to that of wings, and many wing-marker genes are upregulated in both. This evidence supports a wing-co-option scenario for the origin of the treehopper helmet. These results suggest that serial homologues may sometimes diverge evolutionarily through large-scale replacement of, rather than merely tinkering with, their shared patterning network.
Introduction

Understanding the origin of morphological novelty is crucial to explaining the diversity of shapes of organisms in our world. One way to explore the sources of novelty is through investigating developmental biology. Recent years have seen great strides made in understanding the evolution of form through changes in the expression of regulatory genes and the effector genes that they control (Carroll 2005, 2008; Davidson and Erwin 2006; Davidson and Levine 2008; Wagner and Lynch 2008). Such evolutionary developmental research programs have yielded new models of how truly new characters, traits, and body parts may come about in the living world.

One way that morphological novelty may arise via modulation of a genetic regulatory network underlying a trait, meaning that the network of genes expressed in the developing tissue do not change over evolution, but the duration or intensity of expression of some of the genes may change (Gellon and McGinnis 1998). Modulation of gene expression may account for phenotypic evolution such as changes in allometry, as for example growth factors being more active in one tissue relative to another, leading to greater tissue proliferation (Emlen et al. 2006; Refki et al. 2014), which could result in drastic shape change. Another model for the evolution of novelty is via co-option of existing gene regulatory networks. Co-option involves the expression of a set of co-regulated genes in a new developmental context, either in a new tissue or at a new time (Monteiro 2012). Co-option has been implicated in insect novelties such as the mimicry patterns of butterflies (Monteiro 2011; Martin et al. 2014), the hardened elytra of beetles (Tomoyasu et al. 2009), and the grasping structures of male water-strider antennae (Khila et al. 2012).
In the arthropod world, many morphological novelties originate as outgrowths of the body wall, for example beetle horns (Moczek 2006; Emlen et al. 2007; Warren et al. 2014), the protective carapace of water fleas of the genus *Daphnia* (Shiga et al. 2017), mayfly gills (O’Donnell and Jockusch 2010), and perhaps even insect wings (Crampton 1916; Niwa et al. 2010; Tomoyasu 2018). A particularly stunning example of morphological novelty is found in treehoppers, sap-sucking insects of the family Membracidae (Hemiptera). These insects are closely related to leafhoppers (family Cicadellidae) but are distinguished by a body wall outgrowth referred to as a helmet (Deitz and Dietrich 1993; Cryan et al. 2000; Dietrich et al. 2017) ([Fig. 1](#)). Anatomically, a treehopper’s helmet is composed of the dorsal body wall of the insect’s anterior-most thoracic segment, and is properly termed the pronotum (Stegmann 1998; Mikó et al. 2012; Yoshizawa 2012). The membracid pronotum is an evagination of the body wall projecting in three directions—anteriorly, posteriorly, and dorsally—to form a three-dimensional cuticular structure. There are more than 3,300 species of treehoppers worldwide (Cryan et al. 2000, 2004; Dietrich et al. 2017), with helmet structural diversity ranging from a simple posteriorly extended shell that barely covers the insect’s mesonotum to dorsally projected and architecturally complex structures that may be twice as tall as the insect *sans* helmet (Evangelista et al. 2017) (see [Fig. 1B](#)). In the ancestral condition, which is maintained in leafhoppers and other hemipterans, the pronotum is a flat, shield-like part of the exoskeleton that lies flush with the mesonotum (dorsal body wall of the second thoracic segment) (Snodgrass 1927; Govind and Dandy 1970; Matsuda 1970; Mikó et al. 2012) ([Fig. 1A](#)). While the prontal identity of the treehopper helmet is clear, it remains a mystery how the dramatic transformation of flat body wall into a complex and often elaborate three-
dimensional structure came about. What changed in development to give rise to the treehopper helmet?

We consider three main hypotheses for the origin of the treehopper helmet. The first is that modulation in the ancestral gene regulatory network underlying body wall development permitted the outgrowth of the pronotal body wall. Under the modulation hypothesis, the evagination of pronotal body wall that becomes the treehopper helmet would result from the up- or down-regulation of genes that pattern treehopper body wall generally, rather than from extensive modifications ("rewiring") of the genetic regulatory network.

Modulation of pre-existing genetic regulatory networks has recently been implicated in intraspecific variation of *Drosophila* trichome development (Arif et al. 2013) and interspecific variation in *Heliconius* butterfly wing patterns (Martin et al. 2012). Modulation of genes in the insulin/IGF signaling pathway has been shown to underpin the allometry of horn size in the beetle genera *Onthophagus* and *Trypoxylus* (Emlen et al. 2012; Warren et al. 2014).

The other two hypotheses involve co-option, either co-option of part of the leg patterning network or of the wing patterning network. The leg patterning co-option hypothesis arises because treehopper helmets are anatomically similar to beetle horns, in that both are evaginations of body wall. In beetles, the set of genes that combine to generate the proximal-distal axis of legs perform the same function in horns (Moczek and Nagy 2005; Emlen et al. 2006; Moczek and Rose 2009). Specifically, horns of the beetle *Onthophagus* are patterned with a series of nested expression domains of *Distal-less* (*Dll*), *dachshund*, and *homothorax* (*hth*), with *Dll* specifying the distal tip and *hth* the proximal base (Moczek and Rose 2009), a pattern very similar to the proximal-distal axis specification of arthropod legs.
(Lecuit and Cohen 1997; Jockusch et al. 2000). The leg co-option hypothesis for helmet evolution holds that the proximal-distal axis of the treehopper helmet—i.e., the body wall outgrowth—evolved by redeploying a portion of the genetic regulatory network ancestrally involved in leg development. The leg patterning co-option hypothesis for the treehopper helmet is further bolstered by the observation of expression of *Dll* and *hth* in a developing treehopper helmet (*Publilia modesta*) (Prud’Homme et al. 2011).

The wing co-option hypothesis arises from a different interpretation of the data presented by Prud’homme and colleagues (2011), who proposed that the treehopper helmet evolved from the reactivation of an ancient gene regulatory network that patterned wing-like structures on the prothorax of fossil insect taxa, and so the helmet is actually a pair of wings on the normally wingless prothorax. Their conclusion was based in part on the presence of the transcription factor protein Nubbin (Nub) in the developing helmet bud of *Publilia modesta*. In *Drosophila*, the gene *nub* is considered to be a wing selector gene because of its restricted expression domain and its wing-limited function in metamorphosis (Ng et al. 1995; Cifuentes and García-Bellido 1997; Butler et al. 2003), though it has roles in ventral appendage patterning in other insects (Turchyn et al. 2011). While the identity of the helmet as a *bona fide* wing was refuted by Yoshizawa (2012) and Mikó et al. (2012) on morphological grounds, both suggested that wing-gene co-option was a plausible explanation for the similarities between wing and helmet identified by Prud’homme et al. (2011).

Our investigation of the origin of the treehopper helmet deals with theoretical predictions about the divergence of serial homologues—specifically, the notal sclerites of the three thoracic segments (pro-, meso-, and metanotum) Serial homologues are expected to have similar transcriptional profiles because they are built from the same developmental
plan. This expectation has been borne out in studies of flower organs (Chanderbali et al. 2009, 2010) and tetrapod limb development (Liang et al. 2018). However, the co-option of gene regulatory networks should also produce a transcriptional signature, since co-option implies the redeployment of a suite of coregulated genes in a new tissue (Monteiro 2011; Glassford et al. 2015). Thus, each of our three hypotheses lead to predictions of transcriptional similarity between relevant tissues. In the case of the modulation hypothesis, we predict that in both leafhoppers and treehoppers, the pronotum (=treehopper helmet) and mesonotum should be most similar in terms of gene expression (and cluster together in a character tree), because they are serial homologues. In the case of the two co-option hypotheses, we predict that while leafhopper pronotum and mesonotum should still cluster together, the treehopper helmet will cluster with the leg in the case of leg patterning co-option, or with the wings in the case of wing patterning co-option. These predictions can be modeled as simple dendrograms, or character trees, depicting which tissues are most similar to each other. Schematics for the predicted character trees under each hypothesis are summarized in Fig. 2.

We tested these hypotheses by using RNA-seq to compare gene expression among different tissues in two species: a leafhopper, *Homalodisca vitripennis* (Germar 1821), and a treehopper, *Entylia carinata* (Forster 1771). Our results support the wing-patterning co-option hypothesis for the evolution of the treehopper helmet. While leafhopper tissue transcriptomes match the predictions of the ancestral condition, the expression profile of the treehopper helmet is most similar to that of treehopper wings, and several genes in the canonical wing-patterning pathway are upregulated in both tissue types. We also find a strong signal of metamorphosis that shows up in the gene expression profiles of the adult
tissues we sequenced, wings and genitalia. As other comparative transcriptome studies have reported (Plaza et al. 2014; Sudmant et al. 2015), our results also uncover a tissue-specific signal in gene expression comparisons between treehoppers and leafhoppers after controlling for species.

Materials and Methods

Experimental design

For a treehopper species, we chose *E. carinata*, the camel-back treehopper. This species is common in the eastern U.S.A. and Canada, and has the additional beneficial quality of being multivoltine with a short developmental cycle (about two months from when eggs are laid to adulthood.) We chose *H. vitripennis*, the glassy winged sharpshooter, as the leafhopper model. *H. vitripennis* is an emerging model in leafhoppers because it is an important vector of Pierce’s Disease in winery grapes (Kyrkou et al. 2018) and genomic resources are being developed (NCBI BioProject Accession PRJNA168119).

Treehoppers and leafhoppers are hemimetabolous, and like other hemimetabolous insects, their wings, genitalia, and (in the case of treehoppers) helmet are nascent in early instars and acquire their adult form rapidly in the final instar (Mito et al. 2010). Therefore, we dissected 5th (final) instar nymphs to acquire our tissue samples.

We selected eight tissues for transcriptional profiling: eye, pronotum/helmet, mesonotum, second thoracic segment (T2) leg, forewing pads (T2 wings), hind wing pads (T3 wings), abdominal tergum, and ovipositors. This set of tissues was selected because it includes the tissues relevant to our hypotheses (mesonotum, wing, and leg), and also includes three sets of serial homologues (ovipositor/leg, pronotum/mesonotum/abdominal tergum,
fore-/hind wings) and a set of conspicuously metamorphic tissues (ovipositor/wings/helmet) (Fig 2A). The eye was included as a presumptive out-group tissue, because it was expected to be very different in gene expression from the other tissues.

*E. carinata* were raised on *Helianthus annuus* (sunflower) in the UConn EEB Research Greenhouse. The colony was established June 2015 from seven females collected from *Lactuca biennis* (tall blue lettuce) and *Cirsium arvense* (Canada thistle) in Windham and Tolland counties in Connecticut. It has been in continuous culture over three years and produces offspring year-round. To minimize sample variation for RNA-seq, egg clutches from single females were isolated before hatching, and 5th instar nymphs were collected into RNAlater (Invitrogen). In order to amass enough tissue for library construction, nymphs were pooled within broods. This step has the potential to cause bioinformatic complications, due to individual variation in sequence and expression (Wolf 2013), so the size of the pools was minimized as much as possible. Our three biological replicates represent three collections of siblings from different broods (raised at different times) with between 6 and 12 individuals comprising each pool. *H. vitripennis* were reared by the California Department of Food and Agriculture Pierce’s Disease Control Program. Single 5th instar nymphs were collected into RNAlater after being pierced through the abdomen to permit saturation of tissues. After 24 hours at room temperature, RNAlater preserved specimens were frozen at -20 °C and stored for 1 week to 4 months prior to dissection. *H. vitripennis* nymphs are larger than *E. carinata* nymphs, and so these pools comprised no more than 4 individuals.

*RNA-seq library construction*

Preserved nymphs were dissected under RNAlater. Dissected tissues from each pool were stored in RNAlater at -20 °C until extraction. To mitigate batch processing effects,
tissue pools were assigned random numbers prior to RNA extraction, and this processing queue was used through all following procedures. RNA was extracted with TRIzol reagent (Invitrogen), according to the manufacturer’s instructions, except for a modification in the precipitation step to accommodate salt carryover from RNAlater. Total RNA was quantified by fluorometry on a Qubit with RNA Broad Range Assay reagents (Invitrogen), and a small quantity was run on a 1% non-denaturing agarose gel to provide a rough estimate of RNA integrity. Poly(A) enrichment was performed with Sera-Mag oligo-d(T) paramagnetic beads (GE Healthcare Sciences) according to standard protocols (Lambert and Williamson 1993; Mészáros and Morton 1996), with the exception that two rounds of enrichment were performed on RNA pools in which poly(A) yield from the first round indicated it was necessary. Poly(A) yield, expected to be between 1% and 5% of the input total RNA, was measured by fluorometry using the Qubit RNA High Sensitivity Assay kit (Invitrogen). For verification of ribosomal RNA depletion, a small number of RNA pools were selected for fragment analysis by Agilent Tape Station.

Illumina-platform compatible libraries were constructed using the Stranded RNA-Seq kit (KAPA), with NEBNext adaptors for Illumina (New England Biolabs), following the manufacturers’ instructions except for using half-scale reactions. To control for technical variability, ERCC RNA spike-ins (Ambion) were added at the first stage of library preparation, at appropriate dilutions according to the manufacturer’s directions. Libraries were barcoded with custom unique dual indices (UDI) purchased from the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley in order to detect and discard any reads subject to the index swapping that occurs on Illumina HiSeq 4000 platforms (Costello et al. 2018). Final libraries were assessed via Agilent TapeStation to
determine average fragment size and concentration. Based on the TapeStation fragment analysis, libraries were diluted and pooled together to have a predicted equal share in the final pool based on nanomolarity.

Library size selection and sequencing

Final size selection (350 – 800 bp) on the pooled libraries was done via Pippin prep to remove adapter dimer and other short fragments. Libraries were sequenced across 4 lanes of the HiSeq 4000 at 2x100bp (one hundred base pair paired-end reads). Both Pippin prep size selection and sequencing were performed at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley.

Pre-analysis data preparation

Quality trimming and adaptor removal was performed with Trimmomatic (version 0.36) (Bolger et al. 2014). Beginning read quality was high, so in order to preserve length of reads (and therefore increase coverage), the MAXINFO option was used instead of options that aggressively cut low-quality bases. The ILLUMINAACLIP option was used in two passes, first to find and remove adaptor and primer sequences, and then to remove poly(A/T) sequences at the beginning or end of reads, as these are expected to complicate de novo assembly (Wolf 2013). ERCC spike-in reads were removed by filtering with Bowtie2 2.3.3 (Langmead and Salzberg 2012), as were ribosomal reads by filtering based on a curated set of species-specific ribosomal RNA contigs, identified by BLAST against first-draft de novo assemblies of E. carinata and H. vitripennis transcriptomes (Hart et al. 2018, SRA Project Accession SRP152991). Tests of further filtering, such as attempting to remove contaminating reads using FastQScreen (Babraham bioinformatics) showed no impact on
transcriptome assembly. Decontamination was instead performed at the annotation step described below.

**Assembly of species reference transcriptomes**

From each species, we selected the specimen that had the largest absolute number of reads post-quality control. Reads for each tissue-specific library were combined across the four lanes of sequencing and assembled with Trinity v. 2.5.1 (Grabherr et al. 2011). The software default parameters were used except for increasing max_pairs_distance to 800 to reflect our actual library fragment distributions and setting min_contig length to 400bp. Reads for each assembly were aligned and transcript abundance estimated using the Bowtie (version 1.1.2)/RSEM (version 1.3.0) pipeline script included with Trinity. We anticipated that the draft assemblies would include some number of spurious transcripts derived from assembly artifacts, i.e. chimeric contigs, and would likely include some transcripts derived from contaminant sources such as bacterial endosymbionts, plant material, or fungal spores.

We chose the EnTAP pipeline (Hart et al. 2018) for annotation of our assemblies, which includes a feature that detects and discards these unwanted contigs. Briefly, this pipeline performs expression filtering (0.5 FPKM threshold), reading-frame selection with GeneMarkS-T (beta version) (Tang et al. 2015), annotation via similarity searching (DIAMOND version 0.4.7) (Buchfink et al. 2015), and protein family assignment and gene ontology (GO) term annotation with eggNOG (Huerta-Cepas et al. 2016). For similarity searching, we used the Swiss-Prot database (The Uniprot Consortium 2017), a set of predicted proteins for the related *Nilaparvata lugens* (Hemiptera: Delphacidae) (NCBI BioProject accession PRJNA260223) (Li et al. 2015), and a curated set of *Drosophila*
proteins from UniRef90 (The Uniprot Consortium 2017). Fungal and bacterial contaminants were filtered out by similarity annotation and orthogroup identification.

The resulting eight predicted tissue-specific proteomes (annotated, decontaminated amino acid translations) for each species were concatenated, and clustered at a 90% identity threshold with the USEARCH cluster_fast algorithm (Edgar 2010). This high threshold was selected to retain true splicing isoforms, while collapsing identical proteins across all tissues. The centroid sequences were selected as the most representative proteins of each cluster, and their corresponding nucleotide contigs were selected to create the refined reference assembly for each species.

Read mapping, abundance estimation, and within-species differential expression analysis

A convenient way to evaluate the similarity of different characters is to impose a tree-like structure—a character tree (Musser and Wagner 2015)—based on hierarchical clustering of gene expression data. We used the following workflow (Fig. 3) to construct dendrograms. We analyzed differential expression of tissues within each species using the perl and R scripts distributed with Trinity 2.4 (Grabherr et al. 2011), which are designed to construct hierarchical clustering diagrams. Using these scripts, trimmed/filtered reads were mapped to their corresponding reference assembly using Bowtie (Langmead et al. 2009). The resulting alignments were then processed through the RSEM software package (Li and Dewey 2011), which converts the raw number of aligned reads to an estimated number of transcripts in the library using a method that accounts for transcript length and total number of reads in the library. The count data for all libraries were analyzed for differential expression using DESeq2 (Love et al. 2014), using a p-value cut-off (false discovery rate) of 0.001. The resulting set of differentially expressed transcripts were used to derive a Euclidean distance
matrix for clustering analysis using hclust (R Core stats v 3.5.1)(R Core Team 2016). Branch support (approximately unbiased and bootstrap probability) for the resulting dendrogram was calculated using the R package pvclust (version 2.0-0) (Suzuki and Shimodaira 2015) with 1000 bootstrap replications. On all figures depicting hierarchical clustering of samples, branch support is depicted with one asterisk for AU > 70% (strongly supported) and two asterisks for AU > 90% (very strongly supported).

The treehopper species we selected, *E. carinata*, has obvious structural similarity between the helmet and the forewing, which has a heavily sclerotized costal-subcostal region with punctate pattern. The hind wings do not feature this trait. In order to rule out transcriptional similarity due to the treehopper forewings having exoskeleton-like features, we re-ran the analysis for treehoppers leaving out forewings. See Appendix A for details.

This analysis workflow was also repeated for each species on subsets of transcripts subsets that were isolated based on functional annotation: transcription factor activity (GO:0003700), anatomical structure development (GO:0048856), and signaling (GO:0023052). These subsets criteria were selected *a priori* due to their expected relevance to our hypotheses.

*Gene identity and ontology analysis*

Based on the results of annotation, several sets of genes were selected for further investigation regarding their putative protein product function and gene ontology (GO) terms assigned. These sets were identified through two separate processes. Clusters of transcripts with similar expression profiles were identified by pruning branches from the hierarchically clustered gene tree produced by DESeq2. For this method, we used a script utility provided in the Trinity package, manually_define_clusters.R. The other process for identifying sets of
genes involved parsing the results of within-species sample-to-sample pairwise differential expression analysis (to determine which list of transcripts were upregulated for one tissue type relative to the other tissue types. In this method, we chose a set of tissues—for example, forewings, hind wings, and pronotum—and for each tissue, selected all of the genes that were significantly upregulated (FDR < 0.001) for that tissue relative to each other tissue that was not under consideration. In other words, to examine the genes that were upregulated in wings and pronotum, we looked at the set of genes that were significantly upregulated in all of those three tissues relative to each of the other tissues.

Because our initial results indicated a similarity between treehopper helmets and wings, we also examined the expression patterns for particular genes in the *Drosophila* wing patterning network, and for particular genes known for their role in notum patterning in *Drosophila*. Expression levels for genes of interest so identified were plotted in R.

Gene ontology (GO) term enrichment analysis was performed using the R package GoSeq (version 1.32.0) (Young et al. 2010) from Bioconductor. We curated a GO term background for each of our species based on the gene ontology terms assigned by eggNOG during the EnTAP annotation process. Scripts and steps used for this process are posted to [https://github.com/fishercera/TreehopperSeq](https://github.com/fishercera/TreehopperSeq). Briefly, having identified a set of transcripts that related to some tissue or set of conditions in the experiment (as described above), we fitted the transcript length data to a probability weighting function in order to account for selection biases arising from gene length, and then used the goseq function to calculate over- or underrepresentation for each GO term annotated in our transcript set relative to that term’s representation in the background (the species’ transcriptome as a whole). A p-value of less than 0.005 for the overrepresentation test was considered significant enrichment for that
term. This p-value was selected as an appropriate threshold after running high-replicate simulations of GO term enrichment for randomly selected transcript IDs.

Removal of semantically redundant terms was performed by REVIGO (Supek et al. 2011) and visualizations of the results were created using R scripts generated by REVIGO implementing the R Treemap package (version 2.4-2) (Tennekes and Ellis 2017).

Interspecies comparisons

Interspecific comparative transcriptome analyses are challenging due to a confounding species signal that results in more closely correlated gene expression between the tissues of one species than between the homologous tissues of different species (Sudmant et al. 2015; see Musser and Wagner (2015) and Liang et al. (2018) for approaches to dealing with this issue). Additionally, gene duplication and loss events in the diverging evolution of the two species’ genomes mean that gene expression may not be directly comparable for all genes. To deal with the latter problem, we chose to concentrate our analysis only on single-copy orthologues between our leafhopper and treehopper species, a common approach (Plaza et al. 2014; Zhou et al. 2014; Connahs et al. 2016). We curated this set with OrthoFinder (Emms and Kelly 2015) using the isoform-level proteomes predicted by EnTAP. We filtered the raw counts matrices and TPM-normalized matrices down to only these single copy orthologues using custom R scripts. We calculated a TPM scaling factor to account for the different number of mapped transcripts for each species as described in Musser & Wagner (2015); see Appendix B for details.

To winnow out the transcripts that primarily differed in expression between species rather than between tissues, we used a classification strategy employing a Poisson log-linear discriminant analysis (PLDA) (Witten 2011; Zararsiz et al. 2017) using the R package
The approach was developed to find biological markers in RNA-seq data for cancer or other diseases, but here we have applied it to determine which transcripts were the best species-specific “markers” and then exclude them from further analysis. Our PLDA model identified 977 “species” markers (discrete control parameters: tuneLength=100, method=repeatedcv, number=30, repeats=10000). We used DESeq2 to perform a variance-stabilization transformation on the matrix of TPM for the 6,671 remaining orthologues, and then performed principle components analysis in R using the function prcomp() on the transposed, variance-stabilized TPM matrix. See Appendix C for full details.

Results

Library construction and sequencing results

We constructed 48 libraries from our tissue pools—8 tissues from 3 biological replicates for each of 2 species. After quality assessment and fragment analysis, three libraries were removed from the final sequencing pool due to low quality or library preparation failure. These libraries were the eye, T3 wing, and ovipositor of one of the three leafhopper replicates. The remaining 45 libraries, sequenced across 4 HiSeq 4000 lanes, yielded 1.83 billion total reads.

Quality trimming and filtering results

99% of sequenced reads were retained through quality trimming. Adapter and poly-A/T removal resulted in less than 1% of reads being dropped. Filtering out ERCC spike-ins removed the anticipated <1% of reads from all libraries. Ribosomal filtering indicated a high degree of variability in the success of Poly-A enrichment using Sera-Mag oligo-(dT) beads,
with putative ribosomal reads making up from as little as 5% to as much as 50% of the remaining library. See Table 1 for per-library details.

**Draft assembly, annotation, and reference assembly construction results**

We have previously estimated the size of the transcriptomes for E. carinata and *H. vitripennis* as 52.8 megabases and 58.8 megabases respectively (Hart et al. 2018). In order to achieve a complete transcriptome, it is estimated that average coverage should be 100x, or about 18.6 million 2x100bp reads. For *E. carinata*, the reference pool’s 8 tissue libraries yielded a total of 137,482,784 paired-end reads and for *H. vitripennis* this number was 197,743,284. Each library was assembled individually, and these draft assemblies averaged (35,000 contigs) with an N50 of (1250bp).

After filtering and annotating with EnTAP, the average number of contigs remaining in each library was 18,419, with an average N50 of 2,724 bp. A total of 147,583 predicted protein sequences for *E. carinata* and 147,133 for *H. vitripennis* were clustered at the 90% threshold level to a final set of 18,675 (19,975 isoforms) centroid protein sequences for *E. carinata* and 17,630 (19,126 isoforms) for *H. vitripennis*. Their corresponding nucleotide sequences comprised the reference assemblies for each species, with a final N50 of 2,718 bp for *E. carinata* and 3,193 bp for *H. vitripennis*.

**Differential expression**

Differential expression analysis yielded a total of 3,090 DE features for *E. carinata* and 4,405 for *H. vitripennis* across the full transcriptomes. The subset analyses, performed on a selected set of annotated transcripts in each reference assembly, produced differing numbers of differentially expressed features. In *H. vitripennis*, 968 of the 3,492 anatomical structure development genes, 730 of the 2,228 signaling pathway genes, and 180 of the 713
transcription factor genes were differentially expressed. In *E. carinata*, 615 of the 3,496 anatomical structure development genes, 528 of the 2,300 signaling pathway genes, and 160 of the 791 transcription factor genes were differentially expressed.

*Hierarchical clustering results*

Hierarchical clustering of the tissues based on expression of DE genes shows that in *H. vitripennis*, the pronotum is most similar to the mesonotum, as expected of serial homologues (Fig. 4B). In *E. carinata*, however, the helmet (=pronotum) is most similar to the fore- and hind wings, as we predicted under the wing co-option hypothesis (Fig 4A). Ovipositors in both species clustered with wings as well, though in *Entylia carinata* ovipositors fall outside the helmet/wing cluster. The signature of wing/helmet similarity was robust to an analysis that left out the forewings, ruling out the possibility that the treehopper helmet is similar to treehopper wings because of a co-option of exoskeleton patterning in the wings (see Appendix A). Unexpectedly, the abdominal tergum in both species fell outside the rest of the tissues, including the eyes, our putative outgroup tissue. Possible biological processes underlying this result are discussed below.

Our analyses of subsets of the transcriptomes produced different patterns of gene expression from our overall analyses. For the subset of genes classified as anatomical structure development (Fig. 5B), the treehopper helmet again formed a cluster with wings, while the leafhopper pronotum again clustered with the mesonotum (Fig. 5A). Interestingly, in both treehoppers and leafhoppers, the eye tissue clustered just outside the wing/helmet and wing cluster, while the ovipositor tissue fell into a cluster with legs, mesonotum, and abdomen in both species. This result was only found in the anatomical structure development subset. In the analyses of signaling pathway genes, the leafhopper wings formed a cluster
separate from the pronotum, mesonotum, and T2 legs (Fig. 5C), while the treehopper wings and helmet again clustered together (Fig. 5D) indicating a high degree of transcriptional similarity between the treehopper helmet and wings. When we limited the analysis to only those transcripts annotated as transcription factors, the sample clustering for the treehopper species reproduced a pattern of serial homology (Fig. 5F). Pronotum, mesonotum, and abdominal tergum formed a cluster, T2 and T3 formed a cluster, and ovipositor and legs formed a cluster. The serial homology topology is not fully supported in the case of leafhopper transcription factor-based clustering (Fig. 5E), but fore- and hind wings form a well-supported cluster (bootstrap support > 90) apart subtended by eye and a branch leading to the other tissues, and the three body wall tissue types (abdominal tergum, mesonotum, and pronotum) form a well-supported cluster (bootstrap support 81).

Expression of candidate genes

Several genes that are part of the canonical Drosophila wing patterning network are upregulated in both the developing helmet and wings in the treehopper. These include vestigial (vg), apterous (ap), rotund (rn), four-jointed (fj), serum response factor (srf), grainy head (grh), frizzled (fz), wingless (wg), defective proventriculus (dve), engrailed (en), u-shaped (ush), and miniature (m) (Fristrom et al. 1994; Hidalgo 1994; Montagne et al. 1996; Tomoyasu et al. 2000; Sato and Saigo 2000; Furriols and Bray 2001; Nakagoshi et al. 2002; St. Pierre et al. 2002; Roch et al. 2003; Lee and Adler 2004; Ravisankar et al. 2016; Zhou et al. 2017). The majority of these genes are also upregulated in the leafhopper wings relative to other tissues, but not in the leafhopper pronotum, most notably ap, srf, m, wg, fz, fj, and grh (Fig. 6).
In both leafhoppers and treehoppers, several candidate genes that we investigated due to their known involvement in *Drosophila* body-wall patterning were upregulated in common in the helmet/pronotum and mesonotum (Gómez-Skarmeta et al. 1996; Diez del Corral et al. 1999). The gene *araucan/caupolican*, an orthologue of the inparalogous pair of Iroquois-C locus genes, is upregulated in the mesonotum in addition to the pronotum and wings in both leafhoppers and treehoppers. Another Iro-C gene, *mirror*, is upregulated in leafhopper wings, mesonotum, and pronotum, but in the treehopper, it is dramatically upregulated in the pronotum relative to all three other tissue-types. The gene *vg*, typically considered to be a wing-gene but known to have important roles in body-wall patterning (Clark-Hachtel et al. 2013; Medved et al. 2015; Elias-Neto and Belles 2016), is upregulated not only in leafhopper and treehopper wings, but also in the mesonotum and pronotum of both insects. Surprisingly, we did not detect upregulation of either the gene *nubbin* nor *Distal-less* in treehopper pronotum libraries, though we expected it given the antibody staining results found in a closely related treehopper by Prud’homme and colleagues (2011). Other wing- and body-wall related genes that we investigated did not have divergent expression patterns between the two species; these include *ventral-veins lacking*, *spalt major*, *optomotor-blind*, *pannier*, and *pangolin* (Fig. 6) (de Celis et al. 1995; Butler et al. 2003; Tomoyasu et al. 2005).

**Results of GO term enrichment analyses**

The purpose of GO term enrichment analysis is to summarize and simplify the most important functional aspects that differentiate a given subset of genes from the full set of an organism’s genes (Ashburner et al. 2000; The Gene Ontology Consortium 2017). Having identified that the expression of a certain group of genes is relevant to a condition of interest, we can use the GO terms annotated to that set of genes to determine if there is some
biological process, cellular component, or molecular function that crops up more often than expected relative to the background (Young et al. 2010; Huerta-Cepas et al. 2016). In essence, GO term enrichment gives insight into how a set of genes are different from the overall transcriptome, in terms of gene function.

We used GO terms to characterize the function of the sets of genes that were upregulated in common across groups of tissue. Treehopper and leafhopper wing gene expression yielded very similar GO term enrichment pictures. In the biological process tree, developmental process, regulation of biological process, and cell adhesion are highly significant GO terms (Tables 2, 3). The molecular function GO terms enriched in both include structural constituent of cuticle, binding functions, and transcription factor activity (Tables 2, 3). For cellular component GO terms, extracellular region-related terms are enriched in both treehopper and leafhopper wings. This picture changes dramatically for leafhoppers when the set of GO terms considered is limited to those for genes upregulated in both leafhopper wings and leafhopper pronota (Table 2). This set of genes’ GO terms is not significantly enriched for biological process or cellular components; in molecular function, only structural constituent of cuticle and binding functions remain, but not transcription factor activity. Contra leafhoppers, the set of enriched GO terms for genes upregulated in both treehopper wings and treehopper helmets includes many of the terms related to wings (Table 3). In biological process terms, developmental process and anatomical structure development are significantly enriched; in molecular function, transcription factor activity and protein binding are significantly enriched; and in cellular components, the terms extracellular matrix and extracellular region part are significantly enriched (Table 2).
In both leafhoppers and treehoppers, the abdominal tergum tissue fell outside the rest of the tissues in the overall gene expression hierarchical clustering, indicating that the abdominal tergum gene expression was the most distinct relative to the other tissues. This was a surprising result, because under our ancestral state hypothesis we expected abdominal tergum to cluster with its serial homologues, mesonotum and pronotum. GO term enrichment for terms related to biological processes indicate that in both species’ abdominal tissue, the genes that are significantly upregulated are more likely to be functionally annotated for immune system processes (GO:0002376) and responses to stimulus (GO:0050896), relative to the full transcriptome (Fig. 7).

Interspecies comparisons

OrthoFinder recovered 7,648 single-copy orthologues from our annotated isoform proteomes. Applying a PLDA classifier to the scaled TPM values for these transcripts, we identified 977 transcripts that were sufficient to reliably classify our samples by species, our “species markers”. Even with this step, which is intended to diminish the species signal, the largest amount of variation still appears to arise from differences between species. This species signal precludes an informative hierarchical clustering analysis, because each species would simply fall on its own branch. Therefore, we analyzed transcriptional similarity via principle components analysis (PCA). Principle components 1, 2, and 3 collectively explain 51.2% of the variance in expression for these genes; along these axes, treehopper samples in general fall to the negative end while leafhopper samples fall to the positive end. However, the PCA produces a remarkably similar pattern in tissue clustering between the two species, except for the position of each species’ pronotum tissues relative to other tissues. (Fig. 8A,
B). Treehopper pronota fall near treehopper wings, while leafhopper pronota fall near mesonota and legs.

**Discussion**

*Differential gene expression results support the wing co-option hypothesis for the treehopper helmet*

Because co-option implies the redeployment of a suite of genes in a novel context, we expected to see a signal of co-expressed genes in the novel tissue and the original context tissue. This is what we observed in the differential expression results for the treehopper helmet and the treehopper wings. By contrast, the leafhopper pronotum gene expression is most similar to its serial homologue, correlating the similarity between the treehopper helmet and wings with the evolution of the novel helmet.

Importantly, the co-option of wing-patterning genes to form the treehopper helmet does not imply a change in identity for treehopper pronotum. That is to say, the treehopper helmet is not a wing or wing-like organ. The differential expression results for transcription factor genes demonstrate that the underlying transcriptional landscape that patterns the treehopper pronotum is still most similar to its serial homologue, the mesonotum, indicating that the genes co-opted from wing development are primarily downstream of regulatory genes. This is consistent with the model of co-option in which one or perhaps a few master regulator genes that pattern one organ or character acquire regulatory control over a suite of genes ancestrally involved in forming a different organ (Carroll 2005; Monteiro 2011).

These results hold between species, as well. After correcting for gene expression differences arising from species-specific transcriptome differences, treehopper helmets are
more similar to leafhopper and treehopper wings in terms of gene expression than they are to
leafhopper pronota, their homologous tissues. This implies that, at least for these single-copy
orthologues, the relative expression of genes during development has not changed in these
homologous tissues over the course of the species’s evolutionary divergence except for the
expression underlying the novel treehopper helmet, which has become more wing-like as a
result of co-option.

Based on the expression patterns of some known wing-patterning genes and on the
enrichment of annotated GO terms, we can suggest some candidate pathways and/or
developmental mechanisms involved in the co-option event. Wings (but not the pronotum) in
leafhoppers, and wings and helmets in treehoppers, express sets of genes that are enriched in
GO annotations relating to extracellular matrices and secretions, developmental processes,
and transcription factors. The extracellular component terms may point to a specific aspect of
wing development that was co-opted: the evagination and bilayered adhesion of epithelium
that characterizes wing-blade outgrowth. Bolstering this speculation is the known role of
many of the candidate genes in established the proximal-distal axis of wings (i.e., the
evagination of the wing blade: *vestigial* (Halder et al. 1998), *apterous* (Cohen et al. 1992),
defective proventriculus* (Nakagoshi et al. 2002)) or the adhesion and intervein region
specification of the wing blade (*e.g.*, miniature (Bilousov et al. 2012), serum response factor
(Fristrom et al. 1994; Montagne et al. 1996), mirror (Zecca and Struhl 2002)).

*Gene expression in serial homologues diverges physiologically but is conserved at the level
of transcription regulation factors*

The fact that the dorsal abdominal body wall tissues have very different gene
expression from the pronotum and mesonotum in both species of bugs calls into question the
assumption that serial homologues should express similar genes. Considering the differences in GO term enrichment between mesonotum/pronotum and abdomen, we suggest that serial homologues with divergent physiological functions in adulthood may have similarly divergent gene expression when considering the whole transcriptome. Developmental and morphogenesis genes are not the primary signal in gene expression, the physiological genes are. In the case of the abdominal tergum, the important physiological difference between it and its serially homologous body wall tissues may be related to the feeding ecology of these insects. As many sap-sucking hemipterans do, leafhoppers and treehoppers shelter endosymbiotic microorganisms in bacteriomes in their abdomens (Mao et al. 2017). Genomic sequence from bacteria previously identified as endosymbionts for each of our species were among the contaminant sequences identified during annotation, indicating that at least some of the abdominal dorsal body wall tissue we sequenced was bacteriome tissue. Additionally, treehopper abdominal tergum gene expression was enriched for GO term 0044419, interspecies interaction between organisms, and both species’ sets of abdomen-related genes were enriched for other immune process terms.

It was therefore intriguing to discover that, in both species, gene expression profiles of only transcription factors recover the pattern expected from serial homology. We suggest that future comparative tissue transcriptome studies may benefit from considering comparisons in light of the functional role predicted for genes based on annotation by similarity and protein domains.

Future directions

Our main results are robust to a variety of analytical approaches. However, the limited taxonomic scope of the project—one treehopper species and one leafhopper
species—makes it difficult to apply our results to treehoppers generally. The family Membracidae has two major clades within it (Cryan et al. 2000; Dietrich et al. 2017), and there are clade-specific differences in the helmet structure, particularly in the extent of the backwards projection. Our treehopper is a species in the “New World” clade, which have helmets that completely cover their mesonota. Treehoppers in the “Old World” clade tend towards elaborate dorsally projected helmets with little posterior projection, and the helmet often does not cover the individual’s mesonotum. Efforts are underway to add a species of treehopper from this clade, to ascertain whether the wing-patterning co-option event took place early in the treehopper family evolution, and thus can be attributed to the evolution of the novel helmet, or if instead the co-option of the wing patterning network is associated with later innovations of helmet morphology.

*What do our results mean for how serial homologues may evolve into more divergent characters over time?*

The evolution of the treehopper helmet necessarily involves the evolution of divergence in form between two serial homologues. Our data suggest a mechanism for the divergence of serial homologues in which novelty arises by the recruitment of a network of genes into the development of one of a set of serial homologues, thus individuating that organ dramatically over a relatively short evolutionary time span rather than tinkering and tuning development through fine-scale modulation. This means that comparative RNA-seq and character tree analysis of other novel organs in other taxa may reveal unexpected similarities between organ types not because of serial homology, but because of co-option.
Acknowledgments

We wish to acknowledge the kind assistance of Youngsoo Son, of the California Department of Food and Agriculture Pierce’s Disease Control Program, for providing *H. vitripennis* specimens.

Author contributions

Anticipated authors of a manuscript based on this chapter are C. R. Fisher, Jill Wegrzyn, and Elizabeth L. Jockusch. CRF and ELJ conceived the study. CRF, JW, and ELJ designed the experimental and data analysis methods. CRF reared specimens, prepared libraries, analyzed the data and wrote the first draft of the manuscript. ELJ and JW edited the manuscript.

Literature cited


Bilousov, O. O., I. A. Kozeretskia, and V. L. Katanaev. 2012. Role of the gene *Miniature* in


Emlen, D. J., Q. Szafran, L. S. Corley, and I. Dworkin. 2006. Insulin signaling and limb-


Wing patterning genes of *Nilaparvata lugens* identification by transcriptome analysis, and their differential expression profile in wing pads between brachypterous and macropterous morphs. J. Integr. Agric. 14:1796–1807.


### Table 1
Summary of quality trimming and ribosomal RNA removal for the RNA-seq libraries in this study.

<table>
<thead>
<tr>
<th>Library</th>
<th>raw reads</th>
<th>post-trimming</th>
<th>% rRNA</th>
<th>final reads</th>
<th>net</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECEF_Abd</td>
<td>23,227,569</td>
<td>23,003,993</td>
<td>5.15%</td>
<td>21,889,131</td>
<td>94%</td>
</tr>
<tr>
<td>ECEF_Eye</td>
<td>14,994,889</td>
<td>14,853,134</td>
<td>3.85%</td>
<td>14,299,939</td>
<td>95%</td>
</tr>
<tr>
<td>ECEF_Leg</td>
<td>24,848,721</td>
<td>24,687,463</td>
<td>6.65%</td>
<td>23,120,630</td>
<td>93%</td>
</tr>
<tr>
<td>ECEF_Meso</td>
<td>23,233,788</td>
<td>22,848,595</td>
<td>6.81%</td>
<td>21,365,045</td>
<td>92%</td>
</tr>
<tr>
<td>ECEF_Ovi</td>
<td>14,457,588</td>
<td>14,249,614</td>
<td>37.25%</td>
<td>9,230,892</td>
<td>64%</td>
</tr>
<tr>
<td>ECEF_Pro</td>
<td>13,985,336</td>
<td>13,897,289</td>
<td>13.92%</td>
<td>12,055,253</td>
<td>86%</td>
</tr>
<tr>
<td>ECEF_Wing2</td>
<td>20,969,693</td>
<td>20,857,757</td>
<td>6.35%</td>
<td>19,589,285</td>
<td>93%</td>
</tr>
<tr>
<td>ECEF_Wing3</td>
<td>25,991,260</td>
<td>25,857,311</td>
<td>36.87%</td>
<td>16,979,394</td>
<td>65%</td>
</tr>
<tr>
<td>ECFisC_Abd</td>
<td>25,953,896</td>
<td>25,826,880</td>
<td>9.53%</td>
<td>23,452,799</td>
<td>90%</td>
</tr>
<tr>
<td>ECFisC_Eye</td>
<td>25,497,617</td>
<td>25,328,462</td>
<td>6.56%</td>
<td>23,759,195</td>
<td>93%</td>
</tr>
<tr>
<td>ECFisC_Leg</td>
<td>29,336,816</td>
<td>29,090,847</td>
<td>7.94%</td>
<td>26,927,762</td>
<td>92%</td>
</tr>
<tr>
<td>ECFisC_Meso</td>
<td>25,438,630</td>
<td>25,198,786</td>
<td>2.64%</td>
<td>24,564,652</td>
<td>97%</td>
</tr>
<tr>
<td>ECFisC_Ovi</td>
<td>20,200,399</td>
<td>20,039,181</td>
<td>7.29%</td>
<td>18,618,932</td>
<td>92%</td>
</tr>
<tr>
<td>ECFisC_Pro</td>
<td>23,239,495</td>
<td>23,093,393</td>
<td>20.25%</td>
<td>18,416,517</td>
<td>79%</td>
</tr>
<tr>
<td>ECFisC_Wing2</td>
<td>23,239,495</td>
<td>21,715,047</td>
<td>93%</td>
<td>18,416,517</td>
<td>79%</td>
</tr>
<tr>
<td>ECFisC_Wing3</td>
<td>29,919,674</td>
<td>29,711,738</td>
<td>27,903,221</td>
<td>93%</td>
<td>75%</td>
</tr>
<tr>
<td>Ecty4_Abd</td>
<td>24,343,022</td>
<td>24,121,921</td>
<td>3.49%</td>
<td>23,324,342</td>
<td>96%</td>
</tr>
<tr>
<td>Ecty4_Eye</td>
<td>24,604,413</td>
<td>24,343,546</td>
<td>10.89%</td>
<td>21,842,271</td>
<td>89%</td>
</tr>
<tr>
<td>Ecty4_Meso</td>
<td>29,132,415</td>
<td>28,903,710</td>
<td>4.49%</td>
<td>27,677,741</td>
<td>95%</td>
</tr>
<tr>
<td>Ecty4_Ovi</td>
<td>28,566,102</td>
<td>28,340,173</td>
<td>3.99%</td>
<td>27,557,138</td>
<td>96%</td>
</tr>
<tr>
<td>Ecty4_Pro</td>
<td>4,509,102</td>
<td>4,478,157</td>
<td>7.66%</td>
<td>4,146,460</td>
<td>92%</td>
</tr>
<tr>
<td>Ecty4_Wing2</td>
<td>30,816,101</td>
<td>30,586,373</td>
<td>6.30%</td>
<td>28,756,824</td>
<td>93%</td>
</tr>
<tr>
<td>Ecty4_Wing3</td>
<td>9,786,714</td>
<td>9,274,273</td>
<td>52.28%</td>
<td>4,717,893</td>
<td>48%</td>
</tr>
<tr>
<td>HV102_Abd</td>
<td>24,693,802</td>
<td>24,532,330</td>
<td>4.35%</td>
<td>23,495,596</td>
<td>95%</td>
</tr>
<tr>
<td>HV102_Eye</td>
<td>23,436,539</td>
<td>23,261,615</td>
<td>22.77%</td>
<td>18,094,539</td>
<td>77%</td>
</tr>
<tr>
<td>HV102_Leg</td>
<td>35,092,710</td>
<td>34,830,393</td>
<td>1.86%</td>
<td>34,200,771</td>
<td>97%</td>
</tr>
<tr>
<td>HV102_Meso</td>
<td>27,087,280</td>
<td>26,869,934</td>
<td>2.96%</td>
<td>26,216,036</td>
<td>97%</td>
</tr>
<tr>
<td>HV102_Ovi</td>
<td>29,624,512</td>
<td>29,291,924</td>
<td>3.91%</td>
<td>28,178,005</td>
<td>95%</td>
</tr>
<tr>
<td>HV102_Pro</td>
<td>25,103,114</td>
<td>24,646,791</td>
<td>2.21%</td>
<td>24,120,203</td>
<td>96%</td>
</tr>
<tr>
<td>HV102_Wing2</td>
<td>29,392,885</td>
<td>29,026,425</td>
<td>25.29%</td>
<td>21,972,857</td>
<td>75%</td>
</tr>
<tr>
<td>HV102_Wing3</td>
<td>34,172,192</td>
<td>33,847,851</td>
<td>24.79%</td>
<td>25,786,602</td>
<td>75%</td>
</tr>
<tr>
<td>HV3a_Abd</td>
<td>30,971,951</td>
<td>30,489,720</td>
<td>1.50%</td>
<td>30,047,911</td>
<td>97%</td>
</tr>
<tr>
<td>HV3a_Eye</td>
<td>33,676,632</td>
<td>33,463,799</td>
<td>24.70%</td>
<td>25,517,945</td>
<td>76%</td>
</tr>
<tr>
<td>HV3a_Leg</td>
<td>26,043,049</td>
<td>25,852,170</td>
<td>21.59%</td>
<td>20,473,536</td>
<td>79%</td>
</tr>
<tr>
<td>HV3a_Meso</td>
<td>27,732,956</td>
<td>27,256,565</td>
<td>32.54%</td>
<td>18,722,058</td>
<td>68%</td>
</tr>
<tr>
<td>HV3a_Ovi</td>
<td>19,528,212</td>
<td>19,354,241</td>
<td>10.60%</td>
<td>17,358,784</td>
<td>89%</td>
</tr>
<tr>
<td>HV3a_Pro</td>
<td>22,471,128</td>
<td>22,266,942</td>
<td>43.96%</td>
<td>12,838,883</td>
<td>57%</td>
</tr>
<tr>
<td>HV3a_Wing2</td>
<td>14,112,100</td>
<td>13,952,378</td>
<td>1.79%</td>
<td>13,712,923</td>
<td>97%</td>
</tr>
<tr>
<td>HV3a_Wing3</td>
<td>31,372,769</td>
<td>31,083,967</td>
<td>29.13%</td>
<td>22,367,163</td>
<td>71%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>HV7_Abd</td>
<td>29,231,825</td>
<td>29,003,230</td>
<td>3.60%</td>
<td>27,989,467</td>
<td>96%</td>
</tr>
<tr>
<td>HV7_Leg</td>
<td>20,685,457</td>
<td>20,377,174</td>
<td>4.30%</td>
<td>19,516,233</td>
<td>94%</td>
</tr>
<tr>
<td>HV7_Meso</td>
<td>13,816,701</td>
<td>13,715,799</td>
<td>58.58%</td>
<td>6,106,913</td>
<td>44%</td>
</tr>
<tr>
<td>HV7_Pro</td>
<td>22,912,613</td>
<td>22,426,612</td>
<td>48.77%</td>
<td>11,830,188</td>
<td>52%</td>
</tr>
<tr>
<td>HV7_Wing2</td>
<td>16,470,406</td>
<td>16,082,786</td>
<td>36.96%</td>
<td>10,284,694</td>
<td>62%</td>
</tr>
</tbody>
</table>
Table 2 GO term enrichment for *H. vitripennis* for the set of transcripts upregulated in the wings and pronotum, and the set of transcripts upregulated in the wings only. Ontology codes: CC = cellular component, MF = molecular function, BP = biological process. Frequency means the frequency with which the term occurs in the transcript set annotations. Log10 p-value measures how significantly over-represented the term is in the set of transcripts relative to the whole transcriptome.

<table>
<thead>
<tr>
<th>Transcript set</th>
<th>term_ID</th>
<th>description</th>
<th>frequency</th>
<th>log10 p-value</th>
<th>ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>H. vitripennis</em> wings &amp; pro.</td>
<td>GO:0001871</td>
<td>pattern binding</td>
<td>0.13%</td>
<td>-3.8834</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0042302</td>
<td>structural constituent of cuticle</td>
<td>1.26%</td>
<td>-7.2426</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0030246</td>
<td>carbohydrate binding</td>
<td>1.17%</td>
<td>-2.3682</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0031012</td>
<td>extracellular matrix</td>
<td>1.87%</td>
<td>-4.7754</td>
<td>CC</td>
</tr>
<tr>
<td><em>H. vitripennis</em> wings</td>
<td>GO:0044421</td>
<td>extracellular region part</td>
<td>6.65%</td>
<td>-2.8517</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0007155</td>
<td>cell adhesion</td>
<td>2.01%</td>
<td>-2.8386</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0044767</td>
<td>single-organism development process</td>
<td>30.72%</td>
<td>-6.5387</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0050789</td>
<td>regulation of biological process</td>
<td>34.90%</td>
<td>-3.0748</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0044707</td>
<td>single-multicellular organism process</td>
<td>29.71%</td>
<td>-3.9876</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0048856</td>
<td>anatomical structure development</td>
<td>29.52%</td>
<td>-5.4388</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0001871</td>
<td>pattern binding</td>
<td>0.13%</td>
<td>-2.7133</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0003700</td>
<td>transcription factor activity, sequence-specific DNA binding</td>
<td>3.36%</td>
<td>-7.9222</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0042302</td>
<td>structural constituent of cuticle</td>
<td>1.26%</td>
<td>-5.5851</td>
<td>MF</td>
</tr>
</tbody>
</table>
Table 3 GO term enrichment for *E. carinata* for the set of transcripts upregulated in the wings and pronotum, and the set upregulated in wings alone.

<table>
<thead>
<tr>
<th>Transcript set</th>
<th>term_ID</th>
<th>description</th>
<th>frequency</th>
<th>log10 p-value</th>
<th>ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. carinata</em></td>
<td>GO:0031012</td>
<td>extracellular matrix</td>
<td>1.87%</td>
<td>-4.7105</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0044421</td>
<td>extracellular region part</td>
<td>6.65%</td>
<td>-2.332</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0044767</td>
<td>single-organism developmental process</td>
<td>30.72%</td>
<td>-4.4225</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0050789</td>
<td>regulation of biological process</td>
<td>34.90%</td>
<td>-2.2416</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0044707</td>
<td>single-multicellular organism process</td>
<td>29.71%</td>
<td>-3.3088</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0048856</td>
<td>anatomical structure development</td>
<td>29.52%</td>
<td>-3.3373</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0003700</td>
<td>transcription factor activity, sequence-specific DNA binding</td>
<td>3.36%</td>
<td>-6.0669</td>
<td>MF</td>
</tr>
<tr>
<td><em>E. carinata</em></td>
<td>GO:005515</td>
<td>protein binding</td>
<td>20.27%</td>
<td>-4.3922</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0031012</td>
<td>extracellular matrix</td>
<td>1.87%</td>
<td>-8.5437</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0043227</td>
<td>organelle</td>
<td>38.98%</td>
<td>-4.4874</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0044464</td>
<td>cell part</td>
<td>64.71%</td>
<td>-3.9552</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0044421</td>
<td>extracellular region part</td>
<td>6.65%</td>
<td>-4.1917</td>
<td>CC</td>
</tr>
<tr>
<td></td>
<td>GO:0007155</td>
<td>cell adhesion</td>
<td>2.01%</td>
<td>-3.0871</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0009605</td>
<td>response to external stimulus</td>
<td>8.60%</td>
<td>-2.3755</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0048589</td>
<td>developmental growth</td>
<td>3.25%</td>
<td>-3.1008</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0050789</td>
<td>regulation of biological process</td>
<td>34.90%</td>
<td>-11.0504</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0051674</td>
<td>localization of cell</td>
<td>2.95%</td>
<td>-3.7315</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0044707</td>
<td>single-multicellular organism process</td>
<td>29.71%</td>
<td>-13.1404</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0044763</td>
<td>single-organism cellular process</td>
<td>42.86%</td>
<td>-6.0246</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0044767</td>
<td>single-organism developmental process</td>
<td>30.72%</td>
<td>-17.4772</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0003006</td>
<td>developmental process involved in reproduction</td>
<td>7.70%</td>
<td>-3.3401</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0048856</td>
<td>anatomical structure development</td>
<td>29.52%</td>
<td>-13.9632</td>
<td>BP</td>
</tr>
<tr>
<td></td>
<td>GO:0003700</td>
<td>transcription factor activity, sequence-specific DNA binding</td>
<td>3.36%</td>
<td>-21.59</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:005515</td>
<td>protein binding</td>
<td>20.27%</td>
<td>-7.7896</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:0042302</td>
<td>structural constituent of cuticle</td>
<td>1.26%</td>
<td>-5.5229</td>
<td>MF</td>
</tr>
<tr>
<td></td>
<td>GO:1901363</td>
<td>heterocyclic compound binding</td>
<td>25.11%</td>
<td>-4.3025</td>
<td>MF</td>
</tr>
</tbody>
</table>
Figure 1. An example of a plesiomorphic pronotum in a leafhopper, and a sample of pronotal diversity within treehoppers. **A.** Adult specimen of the leafhopper species used in this study, *H. vitripennis*. Yellow dashes outline the mesonotum; white dashes outline the pronotum. **B-D.** Treehopper helmets are three dimensional projections of the pronotum and likely aid in predator defense through disguise, mimicry, and crypsis. In all cases shown here, the insect’s mesonotum is covered by its pronotum. **D.** Adult specimen of *E. carinata*, the treehopper species used in this study. abbreviations: pn = pronotum; mn = mesonotum.

Photo credits clockwise from top left: Michael Schmidt, Creative Commons license cc-by-nc-sa; Pavel Kirillov, Creative Commons license cc-by-sa; Patrick Coin, Creative Commons license cc-by-nc-sa; Kelly Swing, Creative Commons license cc-by-nc-sa
**Figure 2** Experimental design and predictions. **A.** Tissue-type color codes that will be used throughout this manuscript to identify each tissue. **B-D.** Models of the predicted patterns of hierarchical clustering based on differential gene expression for three hypotheses. The position of the pronotum changes based on what tissue it is predicted to be most similar to in terms of gene expression. Dashed-line boxes denote groups of serial homologues in each clustering prediction.
Annotation: EnTAP pipeline

ORF prediction: GeneMarkS-T → Gene identity: Similarity search w/ Diamond → Functional annotation: EggNOG

Assembly refinement

Select annotated ORFs for each library → Cluster amino acid translations at 90% (USEARCH) → Select nucleotide sequences of centroids of each cluster

Differential expression analysis

Within species
Align reads to refined library: Bowtie → Estimate transcript abundance: RSEM → Cluster analysis of differentially expressed genes: DESeq2

Between species
Identify single-copy orthologs: OrthoFinder → Combine and filter transcript counts matrices: R::dplyr → Test for DE genes and analyze clusters: DESeq2

Figure 3 Bioinformatics workflow for annotation, assembly refinement, and differential expression analysis.
Figure 4A. Hierarchical clustering of differentially expressed genes in eight tissue types of nymphal treehoppers. Treehopper helmets (Pro) form a cluster with treehopper wings (T3W, T2W), supporting the wing co-option hypothesis. Numbers by nodes indicate AU bootstrap support. Nodes without numbers have bootstrap support > 90.
Figure 4B. Hierarchical clustering of differentially expressed genes in eight tissue types of nymphal leafhoppers. Leafhopper pronota cluster with their serially homologous tissues, the mesonota, as predicted in the ancestral condition. Numbers by nodes indicate AU bootstrap support; numberless nodes have bootstrap support > 90.
Figure 5 Hierarchical clustering patterns summarized for subsets of genes. A, B. Clustering for genes annotated with GO:0048856-anatomical structure development. A. Leafhopper mesonotum and pronotum are most similar to each other, while B. treehopper helmets are most similar to wings. C, D. Clustering for genes annotated with GO:0023052-signaling. C. Leafhopper mesonotum and pronotum form a cluster with legs, while D. treehopper helmets are most similar to wings. E, F. Clustering based on genes annotated as transcription factors. F. Treehopper topology recovers the predictions of serial homology, and E. the topology for leafhoppers recovers wings clustering together and mesonotum and pronotum clustering together.

Numbers by nodes indicate AU bootstrap support. Nodes without numbers have bootstrap support > 90.
Figure 6 Table depicting the upregulation of candidate developmental genes in tissues of *H. vitripennis* (left) and *E. carinata* (right). Several genes that in leafhoppers are upregulated in wings but not pronotum are upregulated in wings and helmet in treehoppers. The overall pattern suggests a large-scale co-option of wing related developmental genes in *E. carinata*.
Figure 7. The GO terms enriched in the abdominal tergum for leafhoppers (left) and treehoppers (right) suggest that in both species, abdominal body wall is distinct from other body wall tissue because of its close proximity to bacteriomes and endosymbiotic organisms. Size of boxes is relative to the significance (p-value) of statistical over-representation of the GO term. Plot generated in R with scripts generated by ReViGo (Supek et al. 2011) using package TreeMap. Area for each term in the plot is relative to the absolute value of the log of p-value for over-representedness. Larger areas indicate more significantly enriched terms.
Figure 8. Plots of first three principle components of variance-stabilized transcripts-per-million (TPM) for genes not removed by PLDA classifier. Scaling of leafhopper TPM values was performed and variance stabilizing transform was performed on full matrix. **A.** PC1 and PC2. **B.** PC3 and PC1. In both plots, treehopper pronota cluster fall closer to wings than to treehopper mesonota, while leafhopper pronota fall more closely to mesonota. Treehopper samples in general fall more to the negative side of each axis while leafhopper samples fall to the positive side of each axis—a clear demonstration of the species signal. Otherwise, the pattern of distribution for each tissue type is remarkably similar, except for where the pronotum samples fall.
Chapter 2: Understanding the evolution of novelty as changes in regulatory networks via weighted gene co-expression network analysis

Introduction

The evolution of novelty by co-option of gene regulatory networks

One of the ways in which the evolution of novel morphology is thought to occur is by co-option of gene regulatory networks into novel contexts during development. Co-option has often been described as the result of one or a few master regulator genes being redeployed (Tomoyasu et al. 2009; Reed et al. 2011; Monteiro 2012). An example of a novelty which may have arisen via co-option is the dorsal helmet of treehoppers, which is an outgrowth of the first thoracic dorsal body wall (pronotum). Comparative transcriptome analysis has indicated that the treehopper helmet may have evolved by co-option of genes ancestrally involved in wing development (Chapter 1 of this dissertation). Developing treehopper wings and helmets share similar patterns of gene expression, while leafhopper pronota and wings do not. Bolstering this evidence for co-option is the fact that several transcription factors known to be involved in insect wing development are upregulated in the developing treehopper helmet.

The function of many transcription factors has been elucidated through knock-out experiments, characterizing the role of the genes based on the phenotype of null mutants, and this can lead to thinking of certain transcription factors as markers for the genetic regulatory networks they direct. However, transcription factors and other gene products that regulate transcription are highly pleiotropic, and may have functional roles in more than just one or two gene regulatory networks. Genes such as *scalloped (sd)*, *apterous (ap)* and *nubbin (nub)*, which were named for their roles in *Drosophila* wing development (Cohen et al. 1992; Ng et al. 1995;
Halder et al. 1998), are known to have roles in the development of other anatomical structures not only in other arthropods (Damen et al. 2002; Shiga et al. 2017), but in *Drosophila melanogaster* as well. *ap* plays a crucial role in the early central nervous system (Cohen et al. 1992), for example, and the protein Sd, which forms a heterodimer with Vestigial to promise wing growth in *Drosophila* wing imaginal disks, forms a complex with protein Yorkie (Yki) to regulate the Hippo-signaling pathway in eye development (Guo et al. 2013). Therefore, identifying co-option as the causal mechanism by which a novel character evolved should involve an investigation not only of the genes expressed in said character, but the relationship of genes to other genes in genetic regulatory networks.

In order to support the hypothesis that the treehopper helmet evolved by co-option of wing-patterning genes, it would be useful to explore to what extent the networks that are expressed in wings are also expressed in the treehopper helmet. If co-option has occurred, then we should be able to identify a co-expressing module of genes that are expressed in wings of leafhoppers and other related insects, but are expressed in wings and pronotum of treehoppers. Here, we conducted an analysis of gene co-expression networks using our previously generated treehopper and leafhopper gene expression data with wing and pronotum gene expression data for a related insect, the large milkweed bug.

*The promise of co-expression network analysis*

Traditional methods of RNA-Seq gene expression analysis rely on identifying genes that are differentially expressed between two or more conditions, and employing some form of data dimensionality reduction, such as hierarchical clustering or principle components analysis, to draw conclusions based on the similarity of expression between samples (Lamarre et al. 2018). This was the approach we took in Chapter 1 of this dissertation, and it is a powerful analytical
method. Its drawback is that it treats each sample’s gene expression somewhat monolithically, and does not take into account the complexity of how samples differ from each other. A good example of this is the subset analyses of gene expression that we presented in Chapter 1. The treehopper helmet gene expression was more similar to that of wings for the subset of genes related to anatomical structure development or signaling, but was more similar to that of the mesoderm and abdominal tergum for the subset of genes that encoded transcription factors. The abdominal tergum, which we expected to form a cluster with its serial homologue the mesonotum in every analysis, was found in both treehoppers and leafhoppers to be the most divergent tissue from all the rest, except in the analysis of differentially expressed transcription factors. This underscores that different organs are different from each other because of a multitude of processes happening simultaneously, including things like growth and development, metabolism, and functional physiology. The processes are grounded in gene expression and regulation that ought to be separable and distinguishable if we know what groups of genes to look at. Unfortunately, the empirical construction of gene regulatory networks is painstaking work, and only a few regulatory networks have been fully described (Davidson 2009).

Gene co-expression network analysis provides a means to model what the grounding regulatory network might look like without knowing the direct causal interactions of transcription factors, signaling molecules, and other genetic components (Zhang and Horvath 2005; Langfelder and Horvath 2008). Weighted gene co-expression network analysis (WGCNA) is a method of clustering genes based on the correlation of transcripts’ expression with other transcripts. It relies on a property of real-world networks called “scale-free topology,” in which the network has a few highly connected nodes and many nodes that have only one or two connections to other nodes (Zhang and Horvath 2005). The WGCNA method was originally
developed for high-replicate microarray assays, but applies well to RNA-Seq data, if a few constraints are met (Hollender et al. 2014; Brohawn et al. 2016). The primary purpose of WGCNA in the exploration of RNA-Seq data is to uncover the sets of transcripts whose expression patterns change in the same way across different backgrounds. Because these sets of transcripts have highly correlated co-expression, they are likely to be regulated by the same gene regulatory networks.

**Methods**

**Determining species for comparison**

We took a phylogenetic comparative approach to investigate the relationship between gene co-expression networks and body parts (henceforth referred to as “traits” in line with other WGCNA literature), using three species in the order Hemiptera: our model treehopper species, *Entylia carinata*; the leafhopper *Homalodisca vitripennis* (family Cicadellidae) as a close relative lacking the novel focal trait; and the large milkweed bug *Oncopeltus fasciatus* (family Lygaeidae) as a distant relative that also lacks the trait (Fig. 1D). Aside from phylogenetic relatedness, these three species were chosen due to the availability of transcriptome sequence resources and availability of RNA-Seq data for relevant tissues and developmental stages.

**Acquiring RNA-Seq datasets**

For traditional studies of differential expression analysis, we have previously sequenced RNA from multiple tissues of juveniles (fifth instar) of *H. vitripennis* and *E. carinata*, in two separate studies. For the first of these experiments, the tissues studied were mesonotum, pronotum, combined wing buds, and mesothoracic legs; for the second experiment, the same tissues were studied except that fore- and hind wing buds were dissected separately, and eyes, abdominal tergum, and ovipositor tissues were added to the analysis. Abdominal tergum and
ovipositors from the previous specimen sets were also added in the second experiment. Three biological replicates of each species were used in each experiment. The combined data for these experiments comprises 41 libraries for *E. carinata* and 38 libraries for *H. vitripennis* (see Fig. 1). Illumina platform sequencing libraries were constructed as described in Chapter 1 of this dissertation.

*O. fasciatus* is one of the organisms supported by the i5K Initiative (Evans et al. 2013; Poelchau et al. 2015), and was the subject of a recent RNA-Seq study comparing gene expression in developing wings and body wall (Medved et al. 2015). *O. fasciatus* tissue-specific RNA-Seq data were downloaded from the NCBI short-reads archive (accession number SRP066252). Tissue libraries for this study were prepared from pronotum, forewing buds, and hind wing buds of fifth instar nymphs with three biological replicates, along with ectopic pronotal wing buds from three specimens with RNA interference against the gene *Sex-combs reduced*.

*Estimating transcript abundance*

Reference transcriptomes were constructed from full-length transcripts using RSEM (version 1.3.0) (Li and Dewey 2011). For *O. fasciatus*, we constructed a reference transcriptome from the i5K official gene set, full transcripts (oncfas_OGSv1.2_original_transcript.fa), available at the USDA National Agricultural Library site (http://i5k.nal.usda.gov). For *H. vitripennis* and *E. carinata*, the full-length transcripts were refined from the annotated de novo transcriptome assemblies of multiple libraries, as described in Chapter 1 of this dissertation.

Gene expression data were acquired by mapping reads from each of the tissue libraries against that species’ reference transcriptomes with Bowtie (version 1.1.2) (Langmead et al. 2009) and estimating transcript abundance from read counts with RSEM (version 1.3.0) (Li and
Dewey 2011) using the utility scripts packaged with the Trinity (version 2.5.1) transcriptome assembly software (Grabherr et al. 2011). Transcript abundance as TPM (transcripts per million) were normalized with the trimmed means of m-value (TMM) method (Robinson and Oshlack 2010; Lin et al. 2016) to account for differences in library sizes.

*Identifying single-copy orthologs from predicted protein sequences*

For direct comparisons between species, we chose to limit the analysis to identifiable single-copy orthologues. This step is necessary for making one-to-one comparisons, and dramatically cuts down the number of different genes in each assay, which is beneficial for computational resources, but may come at the cost of losing co-expression information useful in network analysis. We used OrthoFinder (version 2.3.1) (Emms and Kelly 2015) to detect single-copy orthologues based on the predicted protein sequences of the transcriptomes. The proteome for *O. fasciatus* was downloaded from the USDA i5k website (oncfas_OGSv1.2_original_peptide.fa). The proteomes for *E. carinata* and *H. vitripennis* were predicted by GeneMarkS-T (Tang et al. 2015) from de novo assemblies as described in Chapter 1. In order to retain a higher number of single-copy orthologues, we first clustered the leafhopper and treehopper proteomes, individually, using USEARCH (method = cluster_fast) (Edgar 2010) with a 60% identity threshold in order to collapse isoforms and partial protein sequences.

For combined analyses, raw TPM values were first scaled between species using a scaling factor (α) described in Musser and Wanger (2015) to account for different numbers of genes in the transcriptomes (see Appendix A of this dissertation), and then size-normalized to account for differences in sequencing depth using the median ratio method (Anders and Huber 2010) as implemented in the R package DESeq2::estimateSizeFactors() function (version 1.20.0) (Love et al. 2014). Size normalization was performed only between libraries of the same species.
After scaling and normalization, the TPM matrices were filtered down to the identified single-copy orthologues and combined for use in gene co-expression network analysis. Because the \textit{O. fasciatus} data set only included pronotum and wings, we created a second TPM matrix with those libraries along with the pronotum and wings libraries from \textit{E. carinata} and \textit{H. vitripennis} (‘Pro and Wings Only’ dataset).

\textit{Constructing and analyzing weighted gene co-expression networks}

We used the R package WGCNA (version 1.64-1) (Langfelder and Horvath 2008) to construct gene co-expression networks for each species, and for three permutations of the combined species set of single-copy orthologues: all samples from all three species, wings and pronota samples only from all three species, and all samples from \textit{H. vitripennis} and \textit{E. carinata}. Using the WGCNA tools, a soft-thresholding power was calculated for each species, which was used to transform the data to better approximate a scale-free network topology (i.e., a network with highly connected hubs) (Zhang and Horvath 2005). An adjacency matrix, which describes the connection strength of each gene to each other gene, was calculated for each species’ transcriptome based on pairwise correlations of gene expression across samples. The topological overlap matrix (TOM) was calculated from the adjacency matrix using the function \texttt{WGCNA::TOMsimilarity()}. Topological overlap is a measure of how similar the network connections are for any two genes; gene A and gene B have high topological overlap if they have similar connection strengths with the same set of other genes. To derive modules, hierarchical clustering of genes using the average linkage method was performed based on a TOM dissimilarity matrix (= 1 – TOM). The resulting dendrogram was dynamically cut (adaptive branch pruning) using the function \texttt{dynamicTreeCut::cutreeDynamic()} with default arguments except that \texttt{deepSplit} was set to 2, \texttt{pamRespectsDendro} was set to \texttt{FALSE}, and minimum cluster
size was set to 15. These modules are arbitrarily assigned names based on the R built-in color palette. Module eigengenes, which are the first principal component of expression for each module, were computed, hierarchically clustered based on their correlation, and used to merge small, similar modules at a tree cut-height of 0.15. Scripts for this analysis were based on the tutorial scripts at

https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/Tutorials/ (Langfelder and Horvath 2016), and are available at

https://github.com/fishercera/TreehopperSeq.

Estimating module-trait significance

WGCNA is often applied to clinical research data (Brohawn et al. 2016; Zheng et al. 2018) which are associated with continuous, quantitative data about traits. In this study, however, we are interested in how the modules detected by WGCNA correlate with tissue types. Therefore, we constructed a binary trait matrix for each analysis we performed, with columns (=traits) for replicate, tissue type, and library, and rows (=sample) for each library. Using binary trait data provides clearer measures of module correlation than using factor data (i.e, one column for tissue types with eight possible values, or factors), and adding correlation tests for all aspects of the experimental design enables us to be confident of real biological signals. Module eigengenes were constructed using the WGCNA function moduleEigengenes() with inputs of the normalized TPM matrix and the previously computed module membership for each gene. Module eigengene correlation to trait was calculated using the Pearson’s correlation method implemented in WGCNA. P-values for module eigengene correlation to trait were adjusted the Benjamini and Hochberg (1995) method to account for multiple comparisons.
WGCNA results often involve dozens of detected modules, and in order to assist in keeping track of modules and their trait relationships, each module is randomly assigned a color derived from R’s built-in color palette. While these names are arbitrary, we continue to use them instead of renaming them based on traits because it is the simplest way to ensure that modules remain fixed through multiple steps of the analysis. Because the results presented here involve five different analyses (one for each of three species, one for the three-species combined analysis for wing and pronotums, and one for *H. vitripennis* and *E. carinata* combined), we refer to the modules below using a prefix denoting which analysis it belongs to: ‘Hvit’ for *H. vitripennis* single species analysis, ‘Ecar’ for *E. carinata*, ‘Ofas’ for *O. fasciatus*, ‘Hemip’ for all samples from all three hemipteran species, ‘WPO’ for wing and pronotum samples only, and ‘Ec_Hv’ for the two species analysis.

*Assigning modules to traits*

We were most interested in modules that could be associated with a tissue type. For some modules this was straightforward, because they were reciprocally the highest correlated module for their highest correlated trait. However, several of our tissue types were strongly correlated with more than one module, and we were interested in whether or not those modules contained a relevant set of genes. In order to deal with sample-specific signals, modules were assigned to tissue types only if the module eigengene was more highly correlated with the tissue type than with any replicate pool or individual sample (with a p-value cut-off of 0.05). An example of such a determination is demonstrated in Fig. 2. Of the *E. carinata* module eigengenes, there were three with reasonably strong positive correlations to the trait ‘Pro’ (corresponding to pronotum libraries). The second and third ranking module eigengenes, ‘Ecar_orangered3’ and ‘Ecar_honeydew1’, were more strongly associated with library ECEF_Pro and ECPRO_C,
respectively. Only the ‘Ecar_lavenderblush3’ module is more strongly associated with trait ‘Pro’ than anything else.

**Characterizing gene co-expression modules**

We characterized gene co-expression modules in five ways: by their module eigengene correlation to traits of interest; by evaluating the relationship between the gene/trait significance with the module membership for the genes belonging to the module; by performing GO term enrichment on the full set of genes belonging to the module; and by identifying the top 20 genes in each module as hub genes, and exploring their annotations. Annotations and GO term tagging were performed via the EnTAP pipeline (Hart et al. 2018) for *E. carinata* and *H. vitripennis*. Annotations for *Oncopeltus fasciatus* were downloaded from i5k (the manual curation set), and were supplemented with a BLASTP search using Diamond (version 0.9.9) (Buchfink et al. 2014) against the Uniprot90 reference protein database (Bateman et al. 2017).

GO term enrichment was performed using the R package goseq from Bioconductor (Young et al. 2010) as described in the methods of Chapter 1 of this dissertation, taking the list of module member genes as the “differentially expressed” set. Gene significance to trait was calculated as the Pearson correlation of each transcript’s expression to the trait of interest. Gene module membership was calculated as the Pearson correlation of each transcript’s expression to the module eigengene’s expression. P-values for gene significance to trait and module membership were calculated with the Student’s t-test as implemented in WGCNA function corPvalueStudent(). The top twenty hub genes were identified based on module membership, where higher module membership implies a higher connectivity of a given gene to other genes in the module. *O. fasciatus* annotations do not include GO terms, so no GO term enrichment analysis was performed.
Results

Ortholog groups and single-copy orthologs identified

The reference transcriptomes for *E. carinata* and *H. vitripennis* contained 19,975 transcripts and 19,126 transcripts, respectively, prior to more stringent clustering (see Chapter 1 of this dissertation). The official gene set for *O. fasciatus* is 19,811 transcripts. Running OrthoFinder on the predicted proteomes for these transcript sets resulted in 4,088 single-copy orthologues. It was expected that the earlier, less aggressive clustering approach (90% identity threshold) had retained splicing isoforms and partial protein sequences that might confound OrthoFinder’s searching, and so 60% identity threshold was selected to force protein isoforms to collapse into a single cluster. This resulted in 15,465 transcripts for *E. carinata* and 14,691 for *H. vitripennis*. Running OrthoFinder on this more rarefied set of transcripts resulted in a total of 7,177 orthogroups with at least one gene from each species, and of that set, 5,084 single-copy orthologues (see Fig. 3). We used this larger dataset for all further combined species analyses.

Modules detected for single species and combined species analyses

WGCNA creates adjacency matrices with an algorithm referred to as soft thresholding, which involves raising the entire expression matrix to a power that permits the detection of correlation networks. The soft thresholding power can be determined empirically, and while it is generally around 6 (Langfelder and Horvath 2016), we calculated it for each dataset we analyzed. For *E. carinata* the calculated soft-threshold power was 5, for *H. vitripennis* it was 6, for *O. fasciatus* it was 8, and for the three-species combined analysis it was 20.

The WGCNA clustering method resulted in 57 modules for *H. vitripennis*, 59 modules for *E. carinata*, and 31 modules for *O. fasciatus*. In the three-species combined analysis with all samples, 16 modules were detected; in the three-species analysis limited to pronota and wings,
33 modules were detected. In the analysis limited to *H. vitripennis* and *E. carinata* single-copy orthologues, 28 modules were detected. Module eigengene and trait correlations are summarized in heatmaps in Figures 4, 6, 8, 10, and 12. Module membership mapped to gene hierarchical clustering dendrograms are shown in Figures 5, 7, 9, and 11.

**Analysis of eye related modules within and across species**

In our RNA-Seq study, an explicit part of the study design was the addition of eye tissue as an anticipated outgroup for the rest of the tissues. In traditional differential expression analyses, this prediction was largely upheld, as eye tissue libraries displayed large blocks (thousands) of upregulated genes that were absent or nearly so in other tissues, and in hierarchical clustering of tissue samples it fell outside all other tissues except for the abdominal tergum. We were interested to see if WGCNA would recover similar sets of genes in *H. vitripennis* and *E. carinata*, and would correlate them with the eye tissue libraries. In *H. vitripennis*, one module, ‘Hvit_turquoise’, was assigned to eyes (2,482 genes, cor=0.93, p=5.49x10^{-12}), and in *E. carinata*, two modules could be assigned to eyes: ‘Ecar_bisque4’ (321 genes, cor=0.99, p=3.97x10^{-31}) and ‘Ecar_green’ (1,558 genes, cor=0.85, p=4.07x10^{-10}). GO term enrichment for these modules is summarized in Fig. 13 and 14. Top 20 hub genes for each module and their annotations and gene-trait significance scores are summarized in Tables 4 and 5. ‘Hvit_turquoise’ and ‘Ecar_bisque4’ include 102 single-copy orthologs between *H. vitripennis* and *E. carinata*, notably *Pax-6*, *orthodenticle*, *rhomboid*, *aristaless*, and genes that code for photoreceptor proteins such as UV opsin and rhodopsin. Two modules were also assigned to eye tissue in the combined *E. carinata* and *H. vitripennis* analysis, ‘Ec_Hv_darkgrey’, (152 genes, cor=0.91, p=4.30x10^{-28}) and ‘Ec_Hv_red’ (210 genes, cor=0.84, p=2.90x10^{-19}). See Figure 15 for a comparison of these modules.
Due to our interest in wing development based on previous investigations of differential expression in treehoppers and leafhoppers, we examined modules assigned to the wing trait in greater detail. For *H. vitripennis*, there were three such modules. In order of correlation score, they were ‘Hvit_darkorange’ (784 genes, cor=0.69, p=1.50x10^{-5}), ‘Hvit_black’ (1,484 genes, cor=0.65, p-value= 0.00081), and ‘Hvit_darkmagenta’ (128 genes corr=0.64, p-value= 1.00x10^{-4}). Scatterplots of gene-trait significance versus module membership are presented in Fig. 16. The gene significance/module membership correlations were ‘Hvit_darkorange’ the lowest at cor=0.54 (p= 1.5x10^{-60}), ‘Hvit_black’ at cor=0.57 (p=1.3x10^{-128}), and ‘Hvit_darkmagenta’ at cor=0.66 (p=2.4x10^{-17}).

GO term enrichment and top-ranked hub genes for these modules indicate their different functions. The ‘Hvit_darkorange’ module is a wing development and metamorphosis module. Its top four over-represented GO terms (p-value < 0.005) are “developmental process”, “anatomical structure development”, “multicellular organismal process”, and “cell adhesion.” These terms are indicative of a module of genes whose annotated functions are related to metamorphosis, particularly of epithelial structures. Supporting this conclusion, its top 20 hub genes (Table 1) include several known to be involved in particular aspects of wing (and generic appendage) development, most notably the genes Notch, Delta, and wing blister (Montagne et al. 1996). The ‘Hvit_black’ module is a DNA transcription module. It has fewer significant GO terms enriched than does ‘Hvit_darkorange’, and those terms are related to metabolic processes and organelle structures. The top 20 ‘Hvit_black’ hub genes appear to be primarily involved in various aspects of transcription and chromatin remodeling (Table 3). The module ‘Hvit_darkmagenta’ is a translation and protein production module. For ‘Hvit_darkmagenta’ there are only 10 significant
enriched GO terms, and by far the most enriched is “structural constituent of ribosome.” (Figure 17B). Its top-ranked hub genes are all various ribosomal proteins (Table 2).

Correlated modules and traits for E. carinata

There was one module assigned to wings in the E. carinata data, and one module assigned to the helmet (=pronotum). The wing module, ‘Ecar_darkolivegreen’, contains 139 genes, and has a correlation with trait ‘Wings’ of 0.51 (p=0.024). The correlation of gene significance for trait versus module membership was 0.16 (p= 0.06, n.s.) (Fig. 18). GO term enrichment for ‘Ecar_darkolivegreen’ is summarized in Fig. 19; highly enriched GO terms include “developmental process” and “anatomical structure development”. Hub genes include a few genes annotated as wing-development related, such as fringe, Echinoid, and tyrosine-protein kinase Src42A (Table 7).

The helmet-related module, ‘Ecar_lavenderblush3’, had 55 genes and a trait correlation score of 0.79 (p=7.14x10^-8). The correlation of module membership to gene trait significance was 0.55 (p= 1.4x10^-5) (Fig. 20A). ‘Ecar_lavenderblush3’ is significantly enriched for a small number of GO terms (Fig. 20B), and the top GO terms are very similar to the GO terms enriched for genes differentially expressed in the helmet (see Chapter 1 of this dissertation), and include “DNA binding transcription factor activity” and “developmental process.” The top 20 hub genes for this module are presented in Table 6. They include the segmental identity HOX gene Sex combs reduced, several wing-related developmental genes, and genes that we have previously identified as highly upregulated in the treehopper helmet, such as u-shaped, delilah, tailup, and araucan/caupolican.

The scatterplots of module membership to gene-trait significance highlights an important facet of WGCNA and similar clustering methods, which is that the genes that compose the
module are not uniformly highly significant to the trait, even if the module eigengene is highly correlated to the trait. A different way of approaching this fact is to look at the top genes associated with a trait regardless of module membership. Table 9 shows the 24 genes that are associated with the combinatorial trait ‘ProNWings’ with a corrected p-value of less than 0.001, along with the modules they to which they were assigned. Most of these genes belong to the ‘Ecar_blue’ module, which could not be assigned to any trait. A few belong to ‘Ecar_lavenderblush3’, and a few belong to ‘Ecar_darkolivegreen.’ This set of transcripts includes grainyhead, apterous, rotund, Wnt, and frizzled-10, genes that we have identified as co-upregulated in the helmet and wings in other analyses.

Modules detected in gene expression data for O. fasciatus.

In general, the modules identified in O. fasciatus gene expression were most strongly related to individual libraries rather than to tissue types (see Fig. 8). One exception to this may be the modules whose eigengenes are highly correlated with the hind wing, ‘Ofas_midnightblue’ (representing 292 genes) and ‘Ofas_darkolivegreen’ (representing 141 genes). The module eigengene correlation to the trait ‘hind wing’ is 0.72 (p=0.38, n.s.) in both cases, and the corrected p-value falls well outside the significance threshold. However, the ‘hind wing’ correlation is the highest correlation these eigengenes have with any trait, meeting our other criterion of assignability. In the case of ‘Ofas_midnightblue’, however, the module eigengene may not be a very good representation of the expression patterns of the module member genes. While ‘Ofas_midnightblue’ has high mean module membership (0.89, s.d. = 0.21), the correlation between gene significance for trait ‘hind wing’ and module membership is 0.38 (p=1.8x10^{-11}). The scatterplot (Fig. 21A) indicates that there are genes with high module-membership scores which are not highly correlated with the hind wing trait.
‘Ofas_darkolivegreen’, on the other hand, has a correlation of 0.57 (p=1.6x10^{-13}) between module membership and gene significance for trait ‘hind wing’. The scatterplot (Fig. 21B) shows relatively few genes with low trait significance and high module membership; in fact, the gene significance/module membership correlation may be influenced more by the relatively low mean module membership score (0.84, s.d.=0.18). Further, the ‘Ofas_darkolivegreen’ correlation with the hind wing is not reducible to the gene expression of any single hind wing sample; all three hind wing samples (Ofas_T3S1, Ofas_T3S2, and Ofas_T3S3) have much lower correlation to the module eigengene than the hind wing itself, with p > 0.05.

**Correlated modules and traits for combined analyses**

Limiting the analysis to libraries represented in all three species, wings and pronota, still shows a very strong correlation of species to modules. The only tissue type that can be assigned a module is ‘Wings’, which can be assigned to ‘WPO_grey’ (597 genes). However, WGCNA analysis reserves ‘grey’ for genes that cannot be properly assigned to a bin due to low topological overlap with other genes. The scatterplot for gene significance to trait ‘Wings’ vs. module membership in module ‘WPO_grey’ shows a skew towards the lower left corner, indicating low significance and low membership. Nevertheless, the correlation of gene significance to module membership is 0.41 (p=0.23, n.s.), and further investigation of the genes assigned to ‘WPO_grey’ may be warranted. Two other candidate modules, ‘WPO_lightcoral’ and ‘WPO_cyan’, do not have significant correlation scores for gene significance to module membership.

In the analysis conducted for the combined *H. vitripennis* and *E. carinata* libraries (‘Ec_Hv’), modules could be assigned to the tissue traits ‘Abd’, ‘Eye’, ‘Leg’, ‘Meso’, and ‘Wings’. For the sake of comparability to single species analyses, we discuss only the ‘Eye’
related modules, ‘Ec_Hv_darkgrey’ and ‘Ec_Hv_red’. ‘Ec_Hv_darkgrey’ (152 genes, mean module membership = 0.77, s.d. 0.15) has a module eigengene to trait ‘Eye’ correlation of 0.91 (p=6.45x10^{-31}). The correlation of gene trait significance for ‘Eye’ to gene module membership is 0.91 (p=3.2x10^{-59}) (see Fig 22A). The other module, ‘Ec_Hv_red’ contains 210 genes (mean module membership = 0.8, s.d. = 0.096), and its module eigengene has a correlation to ‘Eye’ of 0.84 (p=1.8x10^{-21}). However, the scatterplot for gene-trait significance vs. gene module membership (cor=0.88, p=3.5x10^{-69}) shows a slight downward and rightward skew resembling a crescent, characteristic of modules that are more strongly correlated to a specific library than the tissue type (Fig 22B). Individual libraries were not included in the trait matrix for this analysis, but should be investigated before ‘Ec_Hv_red’ can be assigned as an eye-related module.

Discussion

*WGCNA captures biologically relevant gene modules in single species analyses*

The eye modules for *H. vitripennis* and *E. carinata* provide a very clear demonstration of the ability of WGCNA to detect biologically informative and relevant gene clusters. The GO term enrichment for the top eye modules (‘Hvit_turquoise’ and ‘Hvit_bisque4) conform with expectations for eye-related gene expression, and the top annotated hub genes for each module are known to be involved in eye development. One of the hub genes found in both *H. vitripennis* and *E. carinata* ‘Eye’ modules, *Pax6*, has been shown to have similar roles in eye development across the Metazoa (Gehring and Seimiya 2010; Nfonsam et al. 2012; Yoshida et al. 2014).

The modules detected through WGCNA also provide a finer degree of resolution to the gene expression data than more traditional module detection methods. For example, there are three modules in *H. vitripennis* that are about equivalently correlated with wings, but were clustered by WGCNA into distinct bins. Our investigation shows that these modules are
separately identifiable as wing development related genes (‘Hvit_darkorange’), genes related to transcription (‘Hvit_black’), and genes related to protein synthesis (‘Hvit_darkmagenta’).

Similarly, there are two modules in E. carinata associated with eyes, but GO term enrichment and hub gene identity indicates that ‘Ecar_bisque4’ is related to eye anatomy and development, while ‘Ecar_green’ is related to neural anatomy and neurotransmitter synthesis.

The WGCNA analysis also provided further insight into the developmental similarity between treehopper wings and helmet. The E. carinata module related to the helmet, ‘Ecar_lavenderblush3’, has GO term enrichment similar to that of the wing related module ‘Ecar_darkolivegreen’, and the wing development module of H. vitripennis, ‘Hvit_darkorange’. More importantly, the hub genes of the ‘Ecar_lavenderblush3’ module are annotated as genes highly relevant to known wing-development pathways, particularly in vein formation and cell adhesion.

WGCNA clustering did not detect a module of genes that correlated well (and statistically significantly) with both the pronotum and the wings in E. carinata. Since co-option ought to lead to a set of co-regulated genes being expressed in both the original context and the novel context, we had expected that clustering based on co-expression networks should detect such a module. However, the module eigengene-trait correlation heatmap (Figure 6) for E. carinata does appear to indicate that there are groups of modules that are positively correlated with the pronotum and the wings, though the correlation is not statistically significant. Furthermore, the lack of a tightly correlated ‘wings and helmet’ cluster may simply be indicative of the fact that, co-option aside, the treehopper helmet is a very different anatomical structure from the wings. WGCNA modules are identified by the correlation of genes to each other, not by
the correlation of genes to traits, which is why genes that are highly correlated with a trait do not necessarily belong to the same module (such as those in Table 9).

*WGCNA is sensitive to sample-specific gene but expression less sensitive to batch effects than traditional differential expression analysis*

Our results demonstrate that gene co-expression modules that are influenced by individual libraries may be not only detectable, but quite strong. Our investigation of the three candidate modules for the *E. carinata* helmet indicate that individual sample effects should be ruled out before assigning a module to a particular trait. Sample effects were particularly strong in the analysis of *O. fasciatus* gene expression, to the point of being the dominant signal from the data. This is potentially a problem of small replicate data sets. WGCNA performs best when used with high replication of similar samples, which is typical of clinical micro-array expression studies but less common in RNA-Seq expression studies (Langfelder and Horvath 2008). However, it may also be due to high variation in the *O. fasciatus* dataset that was not appropriately controlled for by data normalization techniques.

We have previously limited our RNA-Seq gene expression analyses to libraries from the same sequencing batches, because attempts to combine them resulted in differential expression results in which the strongest signal was whether libraries belonged to the first or second batch. These technical batch effects are a well-known problem in RNA-Seq studies (Wolf 2013), and are likely caused by differential degradation of mRNA molecules, differences in PCR amplification conditions, and other differences in sample handling (Musser and Wagner 2015). Therefore, it is encouraging that neither the *H. vitripennis* nor *E. carinata* WGCNA results show any module that is most strongly correlated with either of the two separate sequencing runs. In this respect, the WGCNA method lives up to its promise: clustering with WGCNA is a marked
improvement on module detection based on hierarchical clustering of differentially expressed genes because it can detect gene clusters that exhibit the same co-expression behavior across samples, even though the magnitude of expression may vary.

Acknowledgements

Funding for this project was provided by an NSF grant to Elizabeth L. Jockusch.

Author contributions

The intended authors for a manuscript based on this chapter are C. R. Fisher and Elizabeth L. Jockusch. CRF and ELJ conceived the study. CRF performed the bioinformatic data collection, performed the analysis, prepared the figures, and wrote the manuscript. ELJ edited the manuscript.
Literature cited


Bateman, A., M. J. Martin, C. O’Donovan, M. Magrane, E. Alpi, R. Antunes, B. Bely, M.
Giorgi, T. Dogan, F. Fazzini, L. G. Castro, L. Figueira, P. Garmiri, G. Georghiou, D.
Nightingale, B. Palka, K. Pichler, D. Poggioli, S. Pundir, L. Pureza, G. Qi, S. Rosanoff, R.
Saidi, T. Sawford, A. Shypitsyna, E. Speretta, E. Turner, N. Tyagi, V. Volynkin, T.
Bridge, S. Poux, N. Redaschi, L. Aimo, G. ArgoudPuy, A. Auchincloss, K. Axelsen, P.
Bansal, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, E. Boutet, L. Breuza, C.
Casal-Casas, E. De Castro, E. Coudert, B. Cuche, M. Doche, D. Dornevil, S. Duvaud, A.
Estreicher, L. Famiglietti, M. Feuermann, E. Gasteiger, S. Gehant, V. Gerritsen, A. Gos, N.
Gruaz-Gumowski, U. Hinz, C. Hulo, F. Jungo, G. Keller, V. Lara, P. Lemercier, D.
Paesano, I. Pedruzzi, S. Pilbout, M. Pozzato, M. Pruess, C. Rivoire, B. Roechert, M.
Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L.
Garavelli, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q.
Wang, Y. Wang, L. S. Yeh, and J. Zhang. 2017. UniProt: The universal protein

Benjamini, Y. ., and Y. Hochberg. 1995. Controlling the false discovery rate : a practical and
 necrosis factor-mediated inflammation as a major abnormality in ALS spinal cord. PLoS
 One 11:1–25.


 required for imaginal disc development in *Drosophila* encodes a member of the LIM family

 common evolutionary origin for wings, breathing organs, and spinnerets. Curr. Biol.
 12:1711–1716.


Edgar, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics


Evans, J. D., S. J. Brown, K. J. J. Hackett, G. Robinson, S. Richards, D. Lawson, C. Elsik, J.
 Coddington, O. Edwards, S. Emrich, T. Gabaldon, M. Goldsmith, G. Hanes, B. Misof, M.
 Muñoz-Torres, O. Niehuis, A. Papanicolaou, M. Pfrender, M. Poelchau, M. Purcell-
 Miramontes, H. M. Robertson, O. Ryder, D. Tagu, T. Torres, E. Zdobnov, G. Zhang, and X.
 Zhou. 2013. The i5K initiative: Advancing arthropod genomics for knowledge, human


Musser, J. M., and G. P. Wagner. 2015. Character trees from transcriptome data: Origin and


Tomoyasu, Y., Y. Arakane, K. J. Kramer, and R. E. Denell. 2009. Repeated co-options of


**Chapter 2: Tables**

**Table 1** Top 20 hub genes for 'Hvit_darkorange' wing-related module

<table>
<thead>
<tr>
<th>gene name</th>
<th>annotated function</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein O-mannosyltransferase 2</td>
<td>O-mannosylation, cellular component synthesis</td>
<td>Uniprot</td>
</tr>
<tr>
<td>Notch</td>
<td>Secreted signaling, Notch signaling pathway</td>
<td>Blastp similarity, biNOG</td>
</tr>
<tr>
<td>multiple coagulation factor deficiency protein 2</td>
<td>ion binding, extra-cellular</td>
<td>biNOG</td>
</tr>
<tr>
<td>Protein O-mannosyltransferase 1</td>
<td>O-mannosylation, cellular component synthesis</td>
<td>Uniprot</td>
</tr>
<tr>
<td>SH3 domain-containing protein Dlish</td>
<td>Negative regulation of Hippo signaling</td>
<td>Uniprot</td>
</tr>
<tr>
<td>protein meteorin</td>
<td>Wing development and metamorphosis</td>
<td>Blastp similarity, iBeetle-base</td>
</tr>
<tr>
<td>protein bowel</td>
<td>imaginal disc-derived leg joint morphogenesis</td>
<td>biNOG</td>
</tr>
<tr>
<td>transmembrane protein 268-like</td>
<td>unknown transmembrane function</td>
<td>Blastp similarity</td>
</tr>
<tr>
<td>Echinus</td>
<td>protein deubiquitination, involved in retinal apoptosis</td>
<td>Uniprot, QuickGO</td>
</tr>
<tr>
<td>wing blister</td>
<td>focal adhesion, epithelial adhesion</td>
<td>Blastp similarity, Uniprot</td>
</tr>
<tr>
<td>fasciclin-2</td>
<td>cell adhesion</td>
<td>Blastp similarity</td>
</tr>
<tr>
<td>quiver</td>
<td>sleep regulation, reduced transmembrane excitability</td>
<td>Uniprot</td>
</tr>
<tr>
<td>RNA-directed DNA polymerase protein naked cuticle homolog 1</td>
<td>Reverse transcriptase</td>
<td>EggNOG</td>
</tr>
<tr>
<td>disco-interacting protein 2</td>
<td>Wnt signaling</td>
<td>Uniprot</td>
</tr>
<tr>
<td>protein Delta</td>
<td>Interacts with disco, neuronal development, appendage</td>
<td>Uniprot</td>
</tr>
<tr>
<td>uncharacterised</td>
<td>development</td>
<td>Uniprot</td>
</tr>
<tr>
<td>putative tyrosine-protein kinase Wsck</td>
<td>Notch signaling</td>
<td>Blastp similarity</td>
</tr>
<tr>
<td>Echinoid</td>
<td>pleckstrin homology domain containing</td>
<td>Blastp similarity</td>
</tr>
<tr>
<td></td>
<td>pseudo-kinase receptor</td>
<td>Uniprot</td>
</tr>
<tr>
<td></td>
<td>Hippo signaling pathway</td>
<td>EggNOG</td>
</tr>
<tr>
<td>gene name</td>
<td>annotated function</td>
<td>source</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Ribosomal protein L11</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>Ribosomal protein L26</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein L36</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein S9</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>60S ribosomal protein L13</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>Ribosomal protein L4</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>Ribosomal protein L31</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein L28</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein L14</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>60S ribosomal protein L18A</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein L14</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein S6</td>
<td>Proteoglycans in cancer (05205), PI3K-Akt signaling pathway (04151), mTOR signaling</td>
<td>BLASTP similarity, EggNOG</td>
</tr>
<tr>
<td></td>
<td>pathway (04150), HIF-1 signaling pathway (04066), Ribosome (03010), Insulin signaling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pathway (04066), Ribosome (03010), Insulin signaling pathway (04910)</td>
<td></td>
</tr>
<tr>
<td>40S ribosomal protein S6</td>
<td>signaling pathway (04910)</td>
<td>EggNOG</td>
</tr>
<tr>
<td>ribosomal protein L3</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>ribosomal protein</td>
<td>Ribosome (03010)</td>
<td>BLASTP similarity</td>
</tr>
</tbody>
</table>
Table 3 Top 20 hub genes for 'Hvit_black' wing-related module

<table>
<thead>
<tr>
<th>gene name</th>
<th>annotated function</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-methylguanosine nucleotidase</td>
<td>Pyrimidine 5'-nucleotidase (UMPH-1)</td>
<td>Uniprot, BLASTP similarity</td>
</tr>
<tr>
<td>Cleavage and polyadenylation specificity factor subunit 5</td>
<td>mRNA surveillance pathway</td>
<td>Uniprot, BLASTP similarity, KEGG</td>
</tr>
<tr>
<td>serine threonine-protein phosphatase</td>
<td>mRNA surveillance pathway</td>
<td>Uniprot, BLASTP similarity, KEGG</td>
</tr>
<tr>
<td>FACT complex subunit Ssrp1</td>
<td>Component of the FACT complex, nucleosome reorganization</td>
<td>EggNOG, BLASTP similarity</td>
</tr>
<tr>
<td>uncharacterized PIH1-domain containing protein</td>
<td>unknown</td>
<td>InterProScan (PFAM)</td>
</tr>
<tr>
<td>Nucleoporin Nup50</td>
<td>NUP50 (Nucleoporin 50 kDa)</td>
<td>EggNOG, BLASTP similarity</td>
</tr>
<tr>
<td>Protein SET-like</td>
<td>Nucleosome assembly protein (NAP)</td>
<td>EggNOG, BLASTP similarity</td>
</tr>
<tr>
<td>ATP-dependent RNA helicase</td>
<td>ATP-dependent RNA helicase</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>Syncrip</td>
<td>Pre-mRNA splicing.</td>
<td>EggNOG, UniProt, BLASTP similarity</td>
</tr>
<tr>
<td>importin subunit alpha</td>
<td>nuclear import receptor</td>
<td>Uniprot, BLASTP similarity</td>
</tr>
<tr>
<td>nucleoplasmin</td>
<td>Chromatin remodeling</td>
<td>Uniprot, BLASTP similarity</td>
</tr>
<tr>
<td>TFIIH basal transcription factor complex helicase</td>
<td>General transcription and DNA repair factor</td>
<td>UniProt, BLASTP similarity</td>
</tr>
<tr>
<td>membrane-bound transcription factor site-2 protease</td>
<td></td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>unknown CHROMO domain protein</td>
<td>Chromo (CHRromatin Organisation MOdifier) domain</td>
<td>EggNOG</td>
</tr>
<tr>
<td>polyadenylate-binding protein 2</td>
<td>Involved in the 3'-end formation of mRNA precursors – Poly(A) tail</td>
<td>EggNOG, BLASTP similarity</td>
</tr>
<tr>
<td>High mobility group protein 20A</td>
<td>Chromatin structure and dynamics</td>
<td>EggNOG, BLASTP similarity</td>
</tr>
<tr>
<td>m7GpppX diphosphatase</td>
<td>RNA degradation pathway, mRNA decapping enzyme</td>
<td>BLASTP similarity, KEGG</td>
</tr>
<tr>
<td>PIH1 domain-containing protein 1</td>
<td>Inherit from biNOG: PIH1 domain containing 1</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>FK506-binding nuclear protein</td>
<td>Peptidyl-prolyl cis-trans isomerase</td>
<td>BLASTP similarity</td>
</tr>
</tbody>
</table>
### Table 4 'Ecar_green' eye related module, top 20 hub genes

<table>
<thead>
<tr>
<th>gene name</th>
<th>annotated function</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>teneurin-a</td>
<td>neuronal development</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td></td>
<td>microtubule motor protein, carries synaptic vesicles</td>
<td>BLASTP similarity, UniProt, InterProScan</td>
</tr>
<tr>
<td>Uncoordinated-104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sodium leak channel non-selective protein</td>
<td>inorganic ion transport and metabolism</td>
<td>BLASTP similarity, EggNOG</td>
</tr>
<tr>
<td>CACNA2D3</td>
<td>subunit of voltage-dependent calcium channel, photo transduction</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>Na/K transporting ATPase</td>
<td>transmembrane transport subunit</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>GABA type B receptor</td>
<td>neurotransmitter receptor</td>
<td>BLASTP similarity, UniProt, EggNOG</td>
</tr>
<tr>
<td>toloid-like</td>
<td>dorsal-ventral patterning, secreted ligand</td>
<td></td>
</tr>
<tr>
<td>Amontillado</td>
<td>hatching behavior, expressed in brain</td>
<td>Uniprot, BLASTP similarity</td>
</tr>
<tr>
<td>scrt</td>
<td>transcriptional repressor Scratch, zinc-finger, brain-related</td>
<td>Uniprot, InterProScan</td>
</tr>
<tr>
<td>uncharacterised NA/K ATPase</td>
<td>sodium pump subunit</td>
<td>InterProScan, EggNOG</td>
</tr>
<tr>
<td>unknown protein</td>
<td>function unknown</td>
<td>EggNOG</td>
</tr>
<tr>
<td>neurologin-4</td>
<td>neuronal cell-adhesion molecule, carboxylesterase family</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>uncharacterised gene myb-like</td>
<td>zinc-finger transcription factor</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>ADAMTS-like</td>
<td>thrombospondin domain, possibly metalloprotease</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>neurexin-3</td>
<td>cell adhesion molecule with laminin G domain</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>uncharacterised NA/K ATPase</td>
<td>cell-membrane transport</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>DSCAM2 cell adhesion molecule</td>
<td>cell adhesion module</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>uncharacterised transmembrane protein</td>
<td>transmembrane function</td>
<td>InterProScan (PFAM)</td>
</tr>
<tr>
<td>Kv channel interacting protein 2</td>
<td>modulates calcium channel activity</td>
<td>BLASTP similarity, UniProt</td>
</tr>
</tbody>
</table>
Table 5 Top 20 hub genes for 'Ecar_bisque4' eye related module

<table>
<thead>
<tr>
<th>gene name</th>
<th>annotated function</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>androgen-dependent TFPI-regulating protein</td>
<td>uncharacterized transmembrane protein</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>cytochrome b5-like</td>
<td>uncharacterized insect cuticle protein</td>
<td>BLASTP similarity (PFAM)</td>
</tr>
<tr>
<td>unknown</td>
<td>uncharacterized transmembrane protein</td>
<td>InterProScan</td>
</tr>
<tr>
<td>unknown</td>
<td>uncharacterized insect cuticle protein</td>
<td>InterProScan (PFAM)</td>
</tr>
<tr>
<td>unknown</td>
<td>uncharacterized protein with receptor domains</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>unknown</td>
<td>amino acid transporter AVT1B</td>
<td>BLASTP similarity, BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>goliath</td>
<td>zinc finger transcription factor</td>
<td>UniProt</td>
</tr>
<tr>
<td>unknown</td>
<td>Sodium/calcium exchanger protein</td>
<td>InterProScan (PFAM)</td>
</tr>
<tr>
<td>RAS and EF-hand domain protein</td>
<td>small GTPase</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>alpha-tocopherol transfer protein</td>
<td>transfer activity</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>stops</td>
<td>termination of phototransduction</td>
<td>UniProt</td>
</tr>
<tr>
<td>transient receptor potential protein</td>
<td>Phototransduction</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>probable collagenase</td>
<td>collagen formation</td>
<td>InterProScan</td>
</tr>
<tr>
<td>GTP cyclohydrolase</td>
<td>folate biosynthesis</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>Akyrin 2</td>
<td>signal transduction</td>
<td>UniProt</td>
</tr>
<tr>
<td>orthodenticle</td>
<td>homeobox gene required for eye development</td>
<td>UniProt</td>
</tr>
<tr>
<td>rhodopsin</td>
<td>light-absorbing retinal pigment</td>
<td>UniProt, BLASTP similarity</td>
</tr>
<tr>
<td>major facilitator superfamily domain-containing protein 12</td>
<td>unknown function</td>
<td>EggNOG</td>
</tr>
<tr>
<td>octopamine receptor Oamb</td>
<td>neurotransmitter receptor</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>Ankyrin-domain containing protein</td>
<td>signal transduction</td>
<td>InterProScan</td>
</tr>
</tbody>
</table>

84
### Table 6: Top 20 hub genes in 'Ecar_lavenderblush3' helmet related module

<table>
<thead>
<tr>
<th>gene name</th>
<th>annotated function</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>unknown</td>
<td>uncharacterized zinc-finger protein HLH transcription factor; wing-vein specification, cell adhesion mediated by integrin</td>
<td>EggNOG</td>
</tr>
<tr>
<td>delilah</td>
<td>monooxygenase involved in neurotransmitter synthesis</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>MOXD1 homolog</td>
<td>zinc finger transcription factor that mediates expression of pannier, involved in hinge/notum differentiation</td>
<td>UniProt</td>
</tr>
<tr>
<td>u-shaped</td>
<td>HOX gene establishing first thoracic segment identity</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>Sex combs reduced</td>
<td>uncharacterized transmembrane/signaling protein</td>
<td>InterProScan (SMART)</td>
</tr>
<tr>
<td>unknown</td>
<td>wing development, vein-forming genes</td>
<td>BLASTP similarity, FlyBase, UniProt, InterProScan (PMFAM, SMART)</td>
</tr>
<tr>
<td>unknown</td>
<td>uncharacterized HLH transcription factor</td>
<td>BLASTP similarity, InterProScan</td>
</tr>
<tr>
<td>enhancer of split mgamma protein</td>
<td>transcription factor repressor, forms dimers uncharacterized monooxygenase protein</td>
<td>UniProt</td>
</tr>
<tr>
<td>unknown</td>
<td>uncharacterized homeobox domain</td>
<td>InterProScan (PFAM, SMART)</td>
</tr>
<tr>
<td>unknown</td>
<td>zinc finger transcription factor that mediates expression of pannier, involved in hinge/notum differentiation</td>
<td>UniProt</td>
</tr>
<tr>
<td>(isoform 2)</td>
<td>transcription factor required for appendage development</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>rotund alpha-trypsin inhibitor</td>
<td>protease inhibitor LIM/homeodomain transcription factor, notum development</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>twist-related protein</td>
<td>dorso-ventral patterning</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>pancreatic lipase-related protein 2-like</td>
<td>phosphatidylcholine 1-acylhydrolase activity</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
<td>Tools</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>facilitated trehalose transporter</td>
<td>sugar transporter, homeobox domain transcription factor, defective proventriculus, involved in wing, eye, and leg morphogenesis</td>
<td>BLASTP similarity, UniProt, InterProScan (PFAM)</td>
</tr>
<tr>
<td>Sex peptide receptor</td>
<td>receptor involved in reproductive and sleep behavior</td>
<td>BLASTP similarity, UniProt similarity, UniProt</td>
</tr>
</tbody>
</table>
### Table 7 Hub genes, 'Ecar_darkolivegreen', wing related

<table>
<thead>
<tr>
<th>gene name</th>
<th>annotated function</th>
<th>source</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGFR kinase substrate 8 transmembrane protein 184B</td>
<td>Regulates cytoskeleton; signal transduction</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>Echinoid</td>
<td>MAPK signaling pathway</td>
<td>UniProt</td>
</tr>
<tr>
<td>unknown</td>
<td>Hippo-signaling pathway</td>
<td>EggNOG</td>
</tr>
<tr>
<td>ATP-binding cassette subfamily G member 8</td>
<td>Transmembrane transporter</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>unknown</td>
<td>Cell cycle progress</td>
<td>UniProt</td>
</tr>
<tr>
<td>serine/threonine protein kinase D</td>
<td>Signaling transduction</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>unknown</td>
<td>Notch signaling in wing development</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>fringe</td>
<td>Epithelial cell development</td>
<td>UniProt</td>
</tr>
<tr>
<td>enabled</td>
<td>Ion transport</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 2</td>
<td>Epithelial cell-cell adhesion, wing development</td>
<td>BLASTP, UniProt</td>
</tr>
<tr>
<td>tyrosine-protein kinase Src42A</td>
<td>Mediates ubiquitination, protein modification</td>
<td>BLASTP similarity, UniProt</td>
</tr>
<tr>
<td>E3 ubiquitin-protein ligase</td>
<td>transmembrane protein of unknown function</td>
<td></td>
</tr>
<tr>
<td>uncharacterized protein</td>
<td>chondrocyte-derived ezrin-like domain-containing protein</td>
<td>BLASTP similarity</td>
</tr>
<tr>
<td>cdep dolichol kinase-like unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 8 Top 20 hub genes for 'Ecar_darkmagenta' wing related module.

<table>
<thead>
<tr>
<th>gene name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cop9 signalosome complex subunit</td>
<td></td>
</tr>
<tr>
<td>RNA polymerase II</td>
<td></td>
</tr>
<tr>
<td>t-complex protein 1 subunit</td>
<td></td>
</tr>
<tr>
<td>Nucleoporin</td>
<td></td>
</tr>
<tr>
<td>endothelial differentiation-related factor-1</td>
<td></td>
</tr>
<tr>
<td>Nucleoporin</td>
<td></td>
</tr>
<tr>
<td>small nuclear ribonucleoprotein</td>
<td></td>
</tr>
<tr>
<td>unknown Cyclin-domain containing protein</td>
<td></td>
</tr>
<tr>
<td>developmentally related GTP-binding protein</td>
<td></td>
</tr>
<tr>
<td>peptidyl-prolyl cis-trans isomerase-like</td>
<td></td>
</tr>
<tr>
<td>Probable prefoldin subunit 2</td>
<td></td>
</tr>
<tr>
<td>aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase</td>
<td></td>
</tr>
<tr>
<td>Probable prefoldin subunit 5</td>
<td></td>
</tr>
<tr>
<td>ribosome biogenesis protein</td>
<td></td>
</tr>
<tr>
<td>proteasome non-ATPase regulatory subunit 12</td>
<td></td>
</tr>
<tr>
<td>T-complex protein 1 subunit beta</td>
<td></td>
</tr>
<tr>
<td>histone deacetylase complex subunit SAP18</td>
<td></td>
</tr>
<tr>
<td>nuclear transcription factor Y subunit B-4</td>
<td></td>
</tr>
<tr>
<td>pleiotropic regulator 1</td>
<td></td>
</tr>
<tr>
<td>programmed cell death protein 5-like</td>
<td></td>
</tr>
</tbody>
</table>
Table 9. *E. carinata* transcripts significantly correlated with the combinatorial trait 'ProNWings' which includes all pronotum and wing samples.

<table>
<thead>
<tr>
<th>Transcript annotation</th>
<th>Module color</th>
<th>s</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NGFI-A-binding protein homolog</td>
<td>Ecar_darkseagreen4</td>
<td>0.80</td>
<td>3.91E-06</td>
</tr>
<tr>
<td>probable phosphoserine aminotransferase</td>
<td>Ecar_blue</td>
<td>0.79</td>
<td>3.91E-06</td>
</tr>
<tr>
<td>protein grainyhead</td>
<td>Ecar_blue</td>
<td>0.80</td>
<td>3.91E-06</td>
</tr>
<tr>
<td>Zinc finger protein rotund</td>
<td>Ecar_blue</td>
<td>0.77</td>
<td>2.69E-05</td>
</tr>
<tr>
<td>MIP18912p/twz</td>
<td>Ecar_lavenderblush3</td>
<td>0.76</td>
<td>3.88E-05</td>
</tr>
<tr>
<td>Protein Wnt</td>
<td>Ecar_blue</td>
<td>0.75</td>
<td>5.97E-05</td>
</tr>
<tr>
<td>protein grainyhead (isoform)</td>
<td>Ecar_blue</td>
<td>0.74</td>
<td>6.98E-05</td>
</tr>
<tr>
<td>CG2663, isoform A</td>
<td>Ecar_lavenderblush3</td>
<td>0.74</td>
<td>7.20E-05</td>
</tr>
<tr>
<td>Unknown</td>
<td>Ecar_blue</td>
<td>0.74</td>
<td>8.85E-05</td>
</tr>
<tr>
<td>frizzled-10-like</td>
<td>Ecar_blue</td>
<td>0.73</td>
<td>9.14E-05</td>
</tr>
<tr>
<td>Unknown</td>
<td>Ecar_blue</td>
<td>0.73</td>
<td>1.44E-04</td>
</tr>
<tr>
<td>CG6426</td>
<td>Ecar_orange</td>
<td>0.72</td>
<td>1.63E-04</td>
</tr>
<tr>
<td>transient receptor potential channel pyrexia</td>
<td>Ecar_blue</td>
<td>0.72</td>
<td>2.23E-04</td>
</tr>
<tr>
<td>Unknown</td>
<td>Ecar_blue</td>
<td>0.71</td>
<td>3.10E-04</td>
</tr>
<tr>
<td>Protein apterous alpha-tocopherol transfer</td>
<td>Ecar_brown4</td>
<td>0.71</td>
<td>3.29E-04</td>
</tr>
<tr>
<td>protein-like</td>
<td>Ecar_blue</td>
<td>0.70</td>
<td>5.36E-04</td>
</tr>
<tr>
<td>uncharacterized protein</td>
<td>Ecar_blue</td>
<td>0.69</td>
<td>6.02E-04</td>
</tr>
<tr>
<td>uncharacterized protein</td>
<td>Ecar_blue</td>
<td>0.69</td>
<td>7.14E-04</td>
</tr>
<tr>
<td>homeobox protein engrailed-1-B-like</td>
<td>Ecar_blue</td>
<td>0.69</td>
<td>7.14E-04</td>
</tr>
<tr>
<td>sodium/potassium/calcium exchanger 4-like</td>
<td>Ecar_darkolivegreen</td>
<td>0.69</td>
<td>7.14E-04</td>
</tr>
<tr>
<td>uncharacterized kinase protein</td>
<td>Ecar_brown4</td>
<td>0.69</td>
<td>7.14E-04</td>
</tr>
<tr>
<td>uncharacterized protein</td>
<td>darkolivegreen</td>
<td>0.68</td>
<td>8.44E-04</td>
</tr>
</tbody>
</table>
Chapter 2: Figures

A. 

<table>
<thead>
<tr>
<th>E. carinata</th>
<th>abd.-tergum</th>
<th>cephal.</th>
<th>mesothorax</th>
<th>metathorax</th>
<th>forewings</th>
<th>hind wings</th>
<th>combined wings</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC_A</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>EC_B</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>EC_C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>ECEF</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>ECEfscC</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>ECt4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

D. 

<table>
<thead>
<tr>
<th>H. vitripennis</th>
<th>abd.-tergum</th>
<th>cephal.</th>
<th>mesothorax</th>
<th>metathorax</th>
<th>forewings</th>
<th>hind wings</th>
<th>combined wings</th>
</tr>
</thead>
<tbody>
<tr>
<td>HV_A</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>HV_B</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>HV_C</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>HV102</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>HV3a</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>HV7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>total</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

C. 

<table>
<thead>
<tr>
<th>O. fasciatus</th>
<th>pronunum</th>
<th>forewings</th>
<th>hind wings</th>
<th>ectopic wings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ofas_1</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ofas_2</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ofas_3</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ofas_RNAi1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ofas_RNAi2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ofas_RNAi3</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 1. A-C). Tissue x Sample matrices of RNA-Seq data used in this study. Empty white boxes indicate missing data. D). Cladogram depicting evolutionary relationships of the three taxa in this study. *E. carinata*, middle picture, is a treehopper; anterior horn and posterior hump are prominent features of its helmet, a double-layered outgrowth of pronotal body wall that is the defining trait of treehoppers.
**Figure 2.** Example of distinguishing species specific modules from modules associated with traits. Scatterplots of gene-trait significance vs. gene module membership. Left column, plots of gene significance to individual samples; right column, plots of gene significance to the cross-cutting tissue trait ‘Pro’ including all treehopper pronotum samples. ‘Ecar_orangered3’ and ‘Ecar_honeydew1’ show very strong correlation with specific, individual libraries. They are not assigned to the tissue trait ‘Pro.’
**Figure 3.** Venn diagram of orthogroup overlap between the three different species as determined by OrthoFinder.
Figure 4. Module eigengene-trait correlation for *H. vitripennis*. Filled in correlation values indicate which relationships are significant (p < 0.05).
**Figure 5.** Module membership mapped against the gene hierarchical clustering dendrogram for *H. vitripennis*. The top block of colors indicate the modules detected before merging at a cut-height of 0.15 of the module eigengene tree. The bottom block of colors indicates module membership after merging clusters.
**Figure 6.** Module eigengene-trait correlation for *E. carinata*. Filled in correlation values indicate which relationships are significant (p < 0.05).
Figure 7. Gene hierarchical clustering and module membership for *E. carinata*.
Figure 8. Module eigengene to trait correlation for *O. fasciatus*. Filled in correlation scores indicate significance relationships (p < 0.05).
Figure 9. Module membership mapped against the gene hierarchical clustering dendrogram for *O. fasciatus*. The top block of colors indicate the modules detected before merging at a cut-height of 0.15 of the module eigengene tree. The bottom clock of colors indicates module membership after merging clusters.
Figure 10. Module eigengene correlation to traits for the three-species wings and pronotum only analysis. Filled in correlation scores indicate significant relationships (p < 0.05).
**Figure 11.** Gene co-expression modules and gene hierarchical clustering dendrogram for the three species WPO analysis. Note that ‘WPO grey’, the second largest module, is the set of genes that could not be placed in any particular module.
Figure 12. Module eigengene to trait correlation for the two-species *H. vitripennis* and *E. carinata* analysis. Filled in correlation values indicate significant relationships.
Figure 13. GO term enrichment for eye related module ‘Hvit_Turquoise’, where p-value for over-representation $< 0.005$. 
Figure 14. GO term enrichment for eye related modules ‘Ecar_bisque4’ and ‘Ecar_green’, where p-value for over-representation < 0.005.
**Figure 15.** Scatterplots for *H. vitripennis* and *E. carinata* combined analysis, eye related modules ‘EC_HV_darkgrey’ (top) and ‘EC_HV_red’ (bottom).
**Figure 16.** Three wing-related modules for *H. vitripennis*. 
Figure 17. GO term enrichment for genes in wing-related modules for *H. vitripennis*. ‘Hvit_darkorange’ is characterized by growth and development related terms; ‘Hvit_darkmagenta’ by ribosomal and protein production terms; ‘Hvit_black’ by transcription and chromatin remodeling terms.
Figure 18. Gene significance for wings versus module membership in module “Ecar_darkolivegreen”. While this is the reciprocally highest correlated module for wings, gene-trait significance is poor. Correlation between module membership and gene significance is poor, and not significant.
Figure 19. GO term enrichment for ‘Ecar_darkolivegreen’. Significantly over-represented GO terms include terms related to growth and development.
Figure 20. ‘E. carinata lavenderblush’ helmet related module. A. Gene significance for Pronotum versus module membership has positive and significant correlation. B. GO term enrichment includes “transcription factor activity” and “developmental process”.
**Figure 21.** Gene significance for hind wing versus module membership for **A.** `Ofas_midnightblue` and **B.** `Ofas_darkolivegreen` module. `Ofas_midnightblue` has a large number of genes with low significance for hind wing but high module membership. In contrast, `Ofas_darkolivegreen` has some genes that are strongly connected to the module (have high module membership) that do not have high significance for hind wing, and vice versa, but overall correlation is positive and significant.
**Figure 22.** Two eye related modules for the combined analysis of *H. vitripennis* and *E. carinata*. **A.** Gene significance versus module membership for ‘EC_HV_darkgrey.’ Correlation is high and highly significant. **B.** The same plot for module ‘EC_HV_red.’ Correlation is still high, but the plot has a slight downward skew, meaning that overall significance for the eye is slightly lower.
Chapter 3: The evolution of adult form via an ancestral, evolutionarily versatile body wall margin module in *Oncopeltus fasciatus*

**Abstract**

Body plan evolution often occurs through the differentiation of serially homologous body parts, and this is particularly apparent in the evolution of insect body plans. Gene expression manipulation experiments that result in either homeotic transformations or loss of particular body parts have demonstrated that ventral appendages along the anterior-posterior axis of insects are serially homologous. Recently, similar experiments have indicated that portions of dorsal and lateral insect body wall may be serially homologous to wings, providing a new perspective on an old question in insect evolution—how did the signature insect novelty, the insect wing, evolve? In this chapter, we investigate the role of several genes associated with wing and body wall development in a hemimetabolous insect, *Oncopeltus fasciatus*. Our results indicate that genes involved in wing development in *O. fasciatus* play similar roles in the development of adult body wall shape and structure, specifically in the development of double-layered flattened evaginations of the body wall. Overall, our results support the existence of a versatile development module for building structures of adhesed cuticularized epithelium, which may have played a central role in the evolution of wings.

**Introduction**

Serial homology refers to the repetition of body parts along an axis. Contrary to the concept of *special homology* (or phylogenetic homology), which describes the same organ or
structure of different species that has its origin in an organ of a shared ancestor, serial homology is conceived of as the same organ being repeated in the same organism, especially as a series along an axis (Bateson 1892; Wagner 1989). Serially homologous body parts are most readily identified in segmented (metameric) animal body plans, with the repetition of segments along the anterior-posterior axis with similar organs, appendages, and other body parts in each segment. Body plan evolution across the tree of life has occurred via differentiation or elaboration of these serially homologous body parts, as for example the differing types of vertebrae in the mammals. Sometimes this process has given to qualitatively novel characters, such as avian wings from tetrapod forelimbs (Wagner 2014). The phylum Arthropoda provides many exceptional models for the investigation of the evolutionary divergence of serially homologous parts because they have small body sizes, short generation times, are more amenable to experimentation than most vertebrates, and possess a many-segmented body plan organized in tagma.

Body parts which are serially homologous are expected to share developmental processes and mechanisms except insofar as they have diverged from their common ground plan. This is reflected in a term used in classical evolutionary literature, homodynamous, implying that the body parts develop under the same power (Bateson 1892; Crampton 1916; Minelli and Peruffo 1991). This prediction has been borne out in numerous studies of the jointed appendages from which phylum Arthropoda takes its name. Mouthpart appendages, legs, antennae, wings, and genital appendages in insects are all serially homologous, and experimental manipulations have demonstrated that they can be homeotically transformed to the identity of other appendages (Bateson 1892; Jockusch et al. 2004; Angelini and Kaufman 2005; Nagy and Williams 2014). In the case of head segments, homeotic transformations
have supported the conclusion that insect mouthparts and antennae are serially homologous with legs in *Drosophila* (Casares and Mann 2001), *Tribolium* (Angelini et al. 2012; Smith et al. 2014), and the hemimetabolous cricket *Gryllus bimaculatus* (Mito et al. 2008). Depletion of the Hox gene *abd-A* in fireflies of the genus *Photuris* causes the homeotic transformation of genital appendages into legs (Stansbury and Moczek 2014), and similar transformations have been observed in *Drosophila* (Gorfinkiel et al. 1999), leading to the inference that these appendages are leg serial homologues. Additionally, suppression of *abd-A* and *Abd-B* in the development of *Bombyx* (Ueno et al. 1992) and *abd-A* in *Tribolium* larvae (Lewis et al. 2000) results in ectopic abdominal appendages, indicating that each appendage-free segment of the abdomen has tissue competent to produce species-typical (and stage-typical) legs. The inference that this tissue is also serially homologous with legs is in line with conclusions drawn from other homeotic transformations.

However, homeotic transformations have also been demonstrated with body part characteristics that are not generally considered serially homologous. In *Tribolium castaneum* larval RNAi, knockdown of *homothorax* (*hth*) resulted in antenna to leg transformations, but also eye to elytron transformations (Smith et al. 2014), and while antennae and legs are serial homologues, eyes and elytra are not. Similarly, some shared patterning may be the result of co-option events, such as the redeployment of a spiracle patterning network in several species of *Drosophila* to produce the novel posterior lobe (Glassford et al. 2015). Examples such as these indicate that while homeotic transformations and other loss of function/knockdown phenotypes are good evidence to support the prediction that serial homologues should share developmental gene patterning, they are not sufficient evidence alone of a serial homology relationship between two organs. This is not
surprising in light of the modular nature of development and cis-regulatory evolution (Wagner 1996; True and Carroll 2002).

By virtue of enhancer region (cis-regulatory) evolution, regulatory genes can be expressed in new domains or at different times during development, and thus re-instantiate the expression of the suite of genes those regulators control (Monteiro 2011; Koshikawa 2015). Such an event is often called co-option, and may often be the cause of shared genetic development between disparate, non-serially-homologous organs. Co-option has been implicated in the origin of a variety of novel body wall outgrowths in insects. Beetle thoracic and head horns, which are structurally simple evaginations of body wall, are developmentally the product of a redeployment of the proximal/distal axis patterning genes of the legs (Kijimoto et al. 2009; Moczek and Rose 2009; Stansbury and Moczek 2014). The treehopper helmet, a novel and often elaborate outgrowth of the prothoracic dorsal body wall, was hypothesized to be a wing-like dorsal appendage on the grounds of shared developmental gene expression and tissue of origin (Prud’Homme et al. 2011). Subsequent morphological analysis rejected this conclusion (Mikó et al. 2012; Yoshizawa 2012), suggesting that the similarity in gene expression might be the result of a co-option of wing-patterning genes. While the overlap of differentially expressed genes between treehopper wings and treehopper helmets is substantial (Chapter 1), the lack of wing hinges or joints and the fact that the helmet comprises the entire prothoracic notum and pleuron makes the conclusion of serial homology less tenable. Wings are distinct from the notum and pleuron of the segment that bears them, though naturalists have debated just what ancestral tissue was the precursor to wings for well over a century (Crampton 1916; Kukalová-Peck 1983; Niwa et al. 2010).
Arguably one of the most important insect novelties, the wings evolved just once, and are restricted to the second and third thoracic segments in extant insects. Hypotheses for their origin can generally be divided into two groups based on the proposed origin tissue, pleural origin hypotheses (lateral body wall) and tergal origin hypotheses (dorsal body wall) (Snodgrass 1927; Tomoyasu et al. 2017). The pleural origin hypothesis includes the leg branch hypothesis, according to which the dorsal branch of the biramous legs of ancestral insects was the precursor of the wing (Wigglesworth 1973; Kukalová-Peck 1983); and the gill origin hypothesis, which posits that tracheated gills were the precursor to wings (Averof and Cohen 1997; Damen et al. 2002; but see also Crampton 1916 for a discussion of late nineteenth century supporters of the gill hypothesis, such as Gegenbauer). Tergal origin hypotheses such as the paranotal lobe hypothesis suggest instead that the wings arose from lateral outgrowths of the dorsal body wall (the tergum) (Crampton 1916; Snodgrass 1927). Niwa et al. (2010) united these two hypotheses into a combinatorial model by demonstrating that abdominal pleural outgrowths in mayflies and bristletails are pre-patterned by regulatory modules marked by *wingless* (*wg*) and *vestigial* (*vg*) expression, and that these insects’ tergal outgrowths express *wg*, *vg*, and *apterous* (*ap*). In this model, wings are tissues of dual origin, composed of both tergal tissue and pleural tissue.

The dual origin hypothesis implies that there are pleural and tergal body wall regions of non-winged trunk segments in winged insects (in other words, the first thoracic segment and all abdominal segments) that are serially homologous with wings (Clark-Hachtel and Tomoyasu 2016). Functional gene studies in beetles subsequently identified putative wing serial homologues that are affected in tandem with wings by the knockdown of wing-marker genes including *vg*, *ap*, and *disheveled* (which encodes a *wg* ligand) (Clark-Hachtel et al.
2013), and tissues in the abdomen that are transformed into homeotic wing-like organs by depletion of the Hox genes *Ubx* and *abd-A* (Ohde et al. 2013). The inference that these are wing serial homologues is supported not only by their shared developmental patterning with wings, but also by their origin in pleural and tergal tissue (i.e., the tissues of wing origin) in non-winged segments. Experiments utilizing CRISPR gene knockouts in the crustacean *Parhyale hawaiensis* suggest that both pleural and tergal tissues—separately—require *vg*, *nubbin (nub)*, and *ap* for proper development (Clark-Hachtel and Tomoyasu 2017), which has been seen as supporting evidence that both tissues are serially homologous with different parts of the wings. However, many of these canonical “wing patterning” genes (specifically *wg*, *vg*, and *scalloped*) are required for the formation of the carapace in *Daphnia pulex* (Shiga et al. 2017), a flattened evagination of the body wall that arises from a different tissue than wings and is oriented on a different axis, which is taken as evidence that a developmental module for body wall margins exists in arthropods and predates the evolution of wings.

Here, we present results of RNA interference during metamorphosis in the hemimetabolous insect *Oncopeltus fasciatus* for genes in the canonical *Drosophila* wing development pathway (*ap*, *vg*, *nub*) in addition to genes known to be involved in notum patterning (*hth*, *araucan/caupolican (ara/caup)*, *tiptop/teashirt (tio)*, and *tailup (tup)*). The resulting phenotypes show both distinct and shared functions for these genes in wing development and body wall margins. In particular, we demonstrate that in *Oncopeltus*, multiple regions of the body wall that are, like wings, structurally double-layered evaginations of epithelium share significant developmental patterning with wings. Our results support the idea that there is a versatile, evolutionarily conserved developmental module in arthropods that is active throughout the thoracic body wall in the production of
flattened, double-layered sheets of body wall. We argue that, in most cases, these thoracic evaginations are not serially homologous or partially serially homologous with wings, though our results do point to the existence of a prothoracic wing serial homologue of both tergal and pleural origin, the junction of the posterior pronotum and propleuron.

Methods

Candidate gene selection, identification, and sequencing

We chose a set of genes to investigate based on their known involvement in either thoracic body wall or wing development in *Drosophila* and *Tribolium*, and then narrowed down our choices with reference to our gene expression data for treehoppers (Chapter 1). For *Of*’ap, *Of*’ara/caup, *Of*’exd, *Of*’mirr, *Of*’tup, *Of*’tio, and *Of*’vg, we downloaded sequences identified via tBLASTn of the orthologous *Drosophila* proteins against the *Oncopeltus fasciatus* transcripts and genome from the i5k Project Workspace (Poelchau et al. 2015). The most complete mRNA sequence available for each gene was used to design PCR primers using Primer3Plus (Untergasser et al. 2012) (Table 1). For *Of*’hth and *Of*’nub, previously designed primers were used (Angelini et al. (unpublished data); Aspiras et al. 2011).

*Oncopeltus* cDNA was constructed from total RNA from fifth instar nymphs with the QScript Flex cDNA kit (Quanta). PCR amplification was empirically optimized for each primer pair. Amplified fragments were purified with silicon membrane columns and used to transform *E. coli* cells with a TOPO-TA cloning kit (Invitrogen). Inserts were sequenced by cycle sequencing to verify identity.
dsRNA design, construction, and injection

The eRNAi web tool (http://e-rnai.org) was used to design efficient dsRNA constructs (Horn and Boutros 2010) and the exact primers suggested by the tool were ordered with T7 promoter sequence at each 5' end (Table 2). These primers were used to amplify template DNA from the cloned plasmids. The template DNA was used to transcribe dsRNA using the HiScribe T7 RNA synthesis kit (New England BioLabs) according to the manufacturer’s instructions, using the longest recommended incubation time of 16 hours. The dsRNA was purified by phenol:chloroform extraction, resuspended in nuclease-free water (Fisher), and quantified by Nanodrop spectrometry. Integrity and proper annealing were assessed by electrophoresis in a non-denaturing 1% agarose gel. Using a physiological buffer of 0.01 mM NaPO₄, 5 mM KCl, and 0.05% McCormick green food coloring, the dsRNAs were diluted to between 1 and 2 μg/μL. Solutions of dsRNA were injected into Oncopeltus nymphs using pulled glass capillary needles mounted on a micromanipulator. Nymphs were anesthetized by cold-stunning and kept on ice until after injection. Our standard injection protocol was adapted from Aspiras et al. (2011) by applying the consecutive injection method of Chesebro et al. (2009). This involves injecting nymphs twice, once during the 4th instar and again during the 5th instar. Negative control groups of equivalent size were injected with double-stranded GFP RNA concurrently with experimental subjects. For Of’ara/caup, which had high lethality, additional experiments were run in which nymphs were injected only in the 4th instar and only in the 5th instar.

Preservation and scoring of resulting adult specimens

For most of the experiments discussed here, the injected individuals that reached adulthood were stored in 75% ethanol at -20°C to preserve tissues. Half of the adults from the
Of'nub and Of'hth experiments were macerated in 80% acetic acid/20% glycerol at 55°C overnight, removing soft internal tissue to better display the cuticular phenotypes. Regardless of preparation method, individuals were scored by checking each discrete body part and recording whether it was wild type or aberrant. Qualitative scoring was used to further record whether specimens presented the same aberrant phenotypes, e.g., “thin, desclerotized wings” and “retracted supracoxal lobes.” Scoring of experimental specimens was performed at the same times as the scoring of their counterpart negative controls.

Photographs were taken of specimens that presented particularly clear examples of phenotypic syndromes, using a Macropod camera setup for Z-plane focus stacking. Optimal settings were 4x zoom with (lens type), shutter speed 1/60, ISO 400, and flash duration setting of 1/16.

Results

Thoracic morphology and serial homology in wild-type Oncopeltus

This study is primarily concerned with the thoracic dorsal and pleural body wall characters of Oncopeltus that are formed from double layered evaginations. These include the two protergal lobes, the mesoscutellar lobe, the supracoxal lobes of the first and second thoracic segments, and the posterior pleural (epimeral) margins of all three thoracic segments. Terminology is adopted from Govind and Dandy (1970), whose description of the thoracic anatomy of Oncopeltus fasciatus most clearly distinguishes double layered marginal evaginations (i.e. “lobes”) from single layered pleural and tergal body wall; synonyms that have been more recently used (Chesebro et al. 2009; Medved et al. 2015) are included parenthetically for clarity.
The pronotum has a longitudinal midline ridge, and anteriorly there are two pairs of indentations on either side of the ridge behind the collar, which are muscle attachment points (Matsuda 1970). Dorsally, the first thoracic segment bears a posteriorly flattened evagination of the pronotum (sensu Mikó et al. 2012), the posterior protergal lobe (Govind and Dandy 1970). The posterior protergal lobe covers the prescutum and scutum of the mesothorax. The dorsal plate of the pronotum is roughly trapezoidal. Anteriorly, the pronotum fuses with the propleuron and prosternum in a ridge-like fold that encircles the head and abuts the posterior cuticle of the eye stalks, called the collar (Tower 1913) or the anterior protergal lobe (Govind and Dandy 1970). The posterior corners of the protergal lobe are rounded, slightly flattened compared to the domed curve of the rest of the pronotum, and cover the wing hinges. The median posterior margin of the protergal lobe curves slightly ventrally and rests against a small transverse rise on the mesothoracic tergum.

The mesoscutellar lobe (scutellum in modern usage) (Fig. 4) is a triangular evagination of the second thoracic segment body wall. It is normally slightly domed, with a distinct midline ridge that ends in a rounded knob covered with setae. The edge of the mesoscutellar lobe is bounded by a rim called the frenum (Tower 1913) which is continuous with the axillary cords of the forewings. The clavus of each hemelytron rests alongside the mesoscutellar lobe when the wings are folded (Medved et al. 2015). The mesoscutellar lobe covers the small, semi-circular plate of the mesothoracic postscutellum, and partially covers the metathoracic prescutum. While the protergal lobe and mesoscutellar lobe are structurally similar as evaginations of dorsal body wall, it is not necessarily the case that they are serially homologous. The mesoscutellar lobe is only a portion of the mesothoracic tergum, which includes the single-layered mesoprescutum, mesoscutum, and mesopostscutellum. The bulk
of the mesonotum, thus, is single-layered, while the pronotum is primarily composed of evaginated body wall and single-layered in a much smaller portion.

Laterally, the protergal lobe is slightly extended and meets the propleuron in a rounded fold to fuse with the pleural plates. The pleura of the pterothorax (meso- and metathorax) extend dorsally to a ridge, the **epimeral protrusion** or **anterior notal process** (Govind and Dandy 1970), which articulates with the basalare plate (Tower 1913) (**Fig. 1**). The pleural plates of the T1 and T2 segments each have **coxal clefts** formed where evaginations of the **episternum** and **epimeron** (**Fig. 1**), the **supracoxal lobes** (**pleural lobes sensu** Matsuda), meet (Tower 1913; Govind and Dandy 1970; Matsuda 1970) (**Fig. 1**). Above the coxal cleft, the pleuron forms a slightly bulging portion called the omium, on the internal surface of which provides the points of attachment for leg muscles (Tower 1913; Govind and Dandy 1970). The T3 **pleuron** does not have a coxal cleft, but does have an **external scent apparatus** (= **scent groove**) (**Fig. 1**), that some authors have homologized with the **coxal clefts** because in some heteropterans, *Oncopeltus* included, the external scent apparatus is in the form of a groove that opens into the metacoxal cavity (Tower 1913; Moody 1930). The **posterior epimeral lobe** (or the **posterior reduplication sensu** Snodgrass (1927)) of each thoracic segment extends to cover part of the pleuron behind it (**Fig. 1**). The sternum of each thoracic segment is continuous with the pleuron (Govind and Dandy 1970; Matsuda 1970). The coxal cavities of the prothorax are not separated from each other by a sternal plate, but the T2 and T3 coxae are. The sternal plate between the mesothorax and metathorax (**metasternum**) is roughly diamond shaped and ends in a small posteriorly pointed evagination, the metaxiphus. The sternal plate between the pro- and
mesothorax (mesosternum) is a large flat field that comes to a triangular, slightly concave bridge between the mesocoxae.

The wings of milkweed bugs are differentiated in that the forewings are partially membranous distally and leathery (coriaceous) proximally, and because of this partial sclerotization the forewings are called the hemelytra (Fig. 7). The hind wings are slightly smaller, and membranous throughout (Fig. 7). Both pairs of wings are used for flight. The wing hinge areas are membranous, and the forewing axillary sclerites present as three articulating sclerotized plates within the wing hinge. The hind wing also has three axillary sclerites that form the cuticularized portion of the wing hinge.

**RNAi overall results**

Experimental and negative control batch sizes and survival rate are summarized in Table 3, and penetrance of phenotypes is summarized in Table 4. Negative control specimens did not display any of the phenotypes discussed below except for a very low rate of antennal disturbances; this is discussed in more detail in the section concerning head segments. Of’tio and Of’ara/caup experimental populations had unexpectedly high 5th instar mortality in experiments using the consecutive injection protocol, due to delay of ecdysis and/or eclosion failure. For this reason, Of’ara/caup RNAi was repeated at 1.0 μg/μL concentration and injected at the fifth instar only, and these results are reported separately except where indicated.

**RNAi effects on posterior pleural margins and supracoxal lobes**

In adult milkweed bugs, the posterior margin of each pleuron extends to cover part of the segment behind it. This margin is formed from a double-layered evagination that folds back under itself to meet the intersegmental membrane. Our experiments indicate that ap,
nub, vg, hth, ara/caup, and tio are all involved to some extent in the normal patterning of pleural margins. In Of’hth RNAi specimens, the pleural margins were highly reduced both posteriorly and dorsally, with the most obvious reduction occurring in the T3 pleuron, which appeared rounded rather than rectangular (Fig 1B). The T2 pleuron reduction that resulted from RNAi targeting Of’hth resulted in the two protuberances of the apodeme developing much more acute rises, owing to a foreshortening in the anterior-posterior axis (Fig. 1B). While very few specimens injected with dsRNA for Of’tio survived to adulthood, those that did showed reduction of the pleura on all three segments so severe that only the pleural plate above the episternum and epimeron still remained. Reductions of the posterior pleural margins of all three segments were also pronounced in Of’nub RNAi specimens, many of which displayed similar phenotypes to Of’hth specimens in the meso- and metathoracic pleura (Fig. 1C). For specimens injected with dsRNA for Of’ap and Of’vg, pleural margin phenotypes had lower penetrance and were largely restricted to the prothoracic pleuron. The edge of the propleuron was reduced in 36% of adult Of’ap RNAi specimens and 77% of Of’vg RNAi specimens.

The supracoxal lobes of the prothorax and mesothorax were affected by knockdown of Of’hth, Of’nub, and Of’tio (Fig. 2B-D). The phenotypes for all three of these genes are contracted lobes of the epimeron and episternum such that the coxal cleft forms an open acute angle. In the case of Of’tio, the trochantins of the pro- and mesothoracic legs were exposed. These results are consistent with the investigation by Medved and colleagues (2015) in which the open coxal cleft phenotype was observed in RNAi targeting Of’nub and Of’tio.
The external scent apparatus located on the metapleural plate, usually a bright orange raised area with a central groove, was reduced by knockdown of Of’hth (Fig. 1B), but at a lower penetrance than other pleural body wall phenotypes for Of’hth RNAi. Knockdown of Of’tio also appeared to disrupt development of the external scent apparatus, but scoring was complicated by the high lethality of Of’tio RNAi. None of the other genes that are part of this study had any effect on the external scent apparatus.

**RNAi effects on dorsal body wall evaginations**

The pronotum was affected to some extent by RNAi targeting each of the genes investigated in this study. Overall pronotum size was reduced in Of’nub, Of’vg, Of’hth, Of’tup, and Of’ara/caup RNAi specimens. For all of these genes, the size reduction in phenotypic specimens was due to reductions of the posteriorly flattened evagination of the pronotum, and not the single-layered T1 tergum. RNAi targeting Of’ap resulted in the dorsal (but not ventral) collar being reduced and retracted from the head such that the collum (back of head, Tower 1913) was visible (Fig. 3C-D). Of’hth RNAi (Fig. 3B) and some Of’ara/caup RNAi specimens (not shown) also displayed a distinctive retraction of the collar.

Knockdown of Of’hth produced the most striking phenotype observed in the posterior margin of the pronotum. In affected specimens, the pronotum is drastically shortened in the anterior-posterior axis due to the posterior margin curling under itself (Fig. 4D). Of’tup RNAi also resulted in a shorter pronotum with a distinctive central notch and much reduced posterior corners that were missing the flattened portions present in the wild type (Fig. 4C). Phenotypes resulting from Of’ap knockdown also showed a rounding and reduction of the posterior corners of the pronotum, but did not have the midline notch (Fig. 4F). Severe Of’ap RNAi specimens had lost the robust, domed shape of the wild-type pronotum, but only in the
most posterior portion, causing a flattened, shelf-like defect. In the case of Of’nub and Of’vg, the pronotum reduction was less extreme than the propleuron reduction, leading to a distinctive square junction at the distal lateral edge where notum and pleuron meet (Fig. 5).

The scutellum was unaffected by Of’nub RNAi (Fig. 4B). RNAi targeting Of’ap resulted in a slightly smaller scutellum that, rather than having a somewhat domed shape, became slightly convex, such that the posterior tip of the scutellum pointed upwards and the frenum was more visible. In some of these specimens, the distinctive knob at the end of the wild-type scutellum was lost, the tuft of setae was reduced, and the very tip of the scutellum showed a small cleft (Fig. 6A-C). RNAi targeting Of’tup resulted in a scutellum that was much smaller than normal, was rounded, and had an apparent increase in the width of the frenum (Fig. 4C). The tip of the scutellum often appeared crumpled in these specimens.

Knockdown of Of’vg, Of’tio, and Of’hth caused shape change in the scutellum without obvious reduction in size. In the case of Of’vg RNAi, specimens showed an expansion of the lateral edges of the scutellum and loss of the triangular shape (Fig. 4H). Of’vg scutella are more nearly rectangular, but retain the midline point ending in a knob. RNAi targeting Of’tio caused the scutellum to develop with two rounded lobes on either side of the midline point (Fig. 4E). In the most severe Of’tio phenotypic specimen, the scutellum had completely lost its triangular shape and appeared nearly identical to the posterior margin of the pronotum. In Of’hth RNAi specimens, the scutellum was more dome shaped and wider than the wild-type scutellum, though still triangular in overall shape (Fig. 4D).

RNAi effects on wings

The genes ap, nub, and vg are well-established determinants of wing size and shape in Drosophila (Cohen et al. 1992; Ng et al. 1995; Halder et al. 1998) and Tribolium (Tomoyasu
et al. 2009; Clark-Hachtel et al. 2013). Medved et al. (2015) previously reported that both Of’nub and Of’vg are required for proper wing development in Oncopeltus, which our results further support. Of’nub RNAi specimens had significantly reduced fore- and hind wings. The proximal portions of the hemelytra were more severely affected than the membranous distal wing blade (Fig. 7D). Of’vg RNAi specimens also had smaller forewings and hind wings. Distal wing venation was disturbed in Of’vg specimens (Fig. 7F), but Of’nub phenotypic wings had normal distal wing venation. Unexpectedly, knockdown of Of’ara/caup also resulted in shorter wings with disturbed venation, and aberrations of the distal wing blade edge (Fig. 7C).

By contrast, nymphal-stage RNAi for Of’ap did not result in size or shape changes, but wings were dramatically changed in texture; rather than being tough and leathery, the entire forewing was thin and membranous, with a marked reduction of pigmentation (Fig. 7B). Forewings in Of’ap RNAi specimens lacked pigment altogether in the wing veins. In some specimens, a clear window-like portion occurred in the middle of the wing. This phenotype, a depigmented and membranous (rather than coriaceous) central field, also was observed in the hemelytral wing pad of 5th instar nymphs that were injected in their 4th instar. This is similar to results reported in Tribolium (Tomoyasu et al. 2009), in which mild ap phenotypes showed reduced sclerotization of the elytra, with clear membrane showing between elytral ridges. The size and shape of the wings were also unaffected by Of’hth RNAi, but hemelytra were desclerotized and depigmented in phenotypic specimens. A less severe pattern of depigmentation was observed in adult specimens injected with dsOf’tup, in which the normally black intervein regions of the hemelytra were lighter or clear (Fig. 7E).
Wing hinges were affected by RNAi targeting *Of*hth specimens, *Of*tup, *Of*ara/caup, and *Of*tio. In *Of*hth RNAi, the axillary sclerites were reduced or missing in many specimens. In the few adult *Of*tio RNAi specimens, the flexible membranous portion and axillary sclerites of the wing hinge were missing, and the wings were connected directly to the notum. Knockdown of *Of*tio may have also resulted in defects of the wing tracheal system; one *Of*tio RNAi specimen died during eclosion with a large bubble of air-inflated membranous epidermis at the right-hand edge of the scutellum.

**RNAi effects on genitalia**

In adult female specimens, *Of*ara/caup RNAi caused doubling of the ovipositor and subgenital plates, a result of an apparent conversion of the A9 segment to A8 identity (results not shown). This phenotype is similar to the ovipositor phenotype reported by Aspiras et al. (2011) for knockdown of *Of*abd-A, in which the second valvulae were transformed into duplicates of the first valvulae, but for *Of*ara/caup homeotic effects extend to the entire segment including valvifers, and surrounding setae. Knockdown of *Of*hth and *Of*nub also disrupted the development of ovipositor valvulae, but in both cases the resulting phenotype was reduced size and failure to form an interlocking ovipositor (results not shown).

**RNAi effects on head segments and head appendages**

In both *Of*hth and *Of*ara/caup phenotypic specimens, the mandibular plate of the head was enlarged and had a swollen appearance due to a duplication of the labial groove separating the mandibular plate from the maxillary plate (Fig. 3B). This phenotype was reported previously for *Of*hth (Aspiras et al. 2011), and our results indicate that both *Of*hth and *Of*ara/caup are necessary for normal patterning of the mandibular and maxillary region of the head in *Oncopeltus*. 
In 92% Of vg RNAi adult specimens, the distal flagellomeres of the antennae were shortened, missing sensory setae, and distorted in shape from the normal smooth cylindrical flagellomere (Fig 8A-B). This phenotype also occurred sporadically in negative control batches and once in the experimental group for the second fragment of tup. The overall rate of distal (A3) flagellomere phenotype was 3.8% in the negative controls and 2.6% in tup experiments. Independent verification with a non-overlapping dsRNA construct is necessary to confirm this heretofore unreported role for vg in antennal development.

Discussion

The existence of a developmental module that patterns flattened body wall outgrowths

Taken together, our results support the existence of an evolutionary versatile module for flattened evaginations of body wall margins (Niwa et al. 2010; Shiga et al. 2017). In Oncopeltus, this module operates in the supracoxal lobes (=pleural lobes), the posterior epimeral lobes of the pleura, and the posterior and anterior prothoracic tergal lobes. All of these structures are affected to some extent by the knockdown of ap, nub, vg, and hth. More importantly, they are affected in very similar ways; nub, vg, and hth are required for these body wall lobes to develop to their proper size, and ap appears to be necessary for the proper development of the dorsal layer of many of these body parts. That these genes are also involved in the developmental patterning of wings in Oncopeltus is unsurprising, as all four are known to be wing-patterning genes in Drosophila and other insects. It is noteworthy, however, that the functional roles of these genes in Oncopeltus wing development is very similar to their roles in the body wall lobes; nub and vg knockdown result in smaller wings, ap knockdown results in a loss of dorsal layer patterning, and hth results in greatly reduced
wing sclerites. The shared patterning of the wings and the body wall lobes is the result of the wings being partially patterned by this body wall margin module.

Indeed, this module likely predates the origin of wings by hundreds of millions of years, because it appears to operate in body wall lobes across Pancrustacea, if not the arthropod phylum at large. The gene \textit{nub} is expressed in the book gills of the xiphosuran chelicerate \textit{Limulus polyphemus} in a manner consistent with wing development rather than leg development (uniform expression across the outgrowth, rather than annulations at future joints) (Damen et al. 2002). The crustacean \textit{Artemia franciscana} expresses both \textit{nub} and \textit{ap} in the epipodites that develop above its legs (Averof and Cohen 1997). More recently, the role of \textit{vg} and \textit{wg} in the development of the carapace of \textit{Daphnia pulex} has been elucidated (Shiga et al. 2017). These previous crustacean studies focused on investigating expression domains, but the advent of CRISPR has made it possible to investigate the functional role of these genes as well in the malacostracan crustacean \textit{Parhyale hawaiensis}. These three wing genes, \textit{ap}, \textit{nub}, and \textit{vg}, are required for the proper development of four flattened body wall outgrowths in \textit{Parhyale}: the tergal margin, the coxal plate, the basis (i.e., the segment between crustacean coxa and ischium), and the gill (\textit{ap} and \textit{nub} only) (Clark-Hachtel and Tomoyasu 2017). Furthermore, based on their expression patterns, \textit{ap} and \textit{vg} appear to be involved in flattened body wall outgrowths in the apterygote insect \textit{Pededontus unimaculatus} (Archaeognatha) (tergal margins) and the paleopteran insect \textit{Ephoron eophilum} (Ephemeroptera) (tergum and lamellate gills) (Niwa et al. 2010). Importantly, these organs arise on a variety of locations of the dorsal and lateral body wall, and while some of them may be wing homologues, it is not necessarily the case that all of them are. What they
certainly have in common is their structural architecture: flattened, double-layered evaginations of body wall.

**Support for the mesoscutellar lobe as a hemipteran novelty patterned independently from other body wall lobes**

While the scutellar lobe of *Oncopeltus* is structurally similar to the posterior protergal lobe and the other body wall evaginations discussed here, in that it is two fused layers of epithelium, our results do not support the conclusion that it is patterned by the same evolutionarily versatile module discussed above. To the contrary, the genes studied seem to have very different functions in the development of the scutellar lobe. Knockdown of *vg* and *hth* resulted in a larger rather than smaller scutellar lobe, along with a shape change. The gene *nub* had no detectable role in the scutellar lobe’s development. Of the genes discussed above as part of the proposed module for body wall margin outgrowths, only *ap* appears to play a similar role in the scutellar lobe and other outgrowths; the dorsal layer of the lobe is reduced, causing a curled-outwards phenotype. On the other hand, the gene *tup*, which in *Drosophila* appears to specifically pattern the median notum (de Navascues and Modolell 2007), appears to play an important role in the *Oncopeltus* scutellar lobe that is divergent from its role elsewhere; the *tup* RNAi scutellar lobe is reduced, while other lobes and the wings are not reduced, but are depigmented, particularly in their margins. Given that the scutellar lobe in *Oncopeltus* (and heteropterans in general) develops within a nymphal bud alongside the hemelytra, it is tempting to speculate that the scutellar lobe arose by the expansion of expression domains of wing genes that then interacted with the median notum patterning genes.
However, other parts of the body also responded to the knockdown of some of these genes in divergent ways. The knockdown of *vg* affected the distal flagellumere of the antennae, which is cylindrical in the wild type rather than flattened. The ovipositors of female *Oncopeltus* showed disruption and even homeotic transformations in response to knockdown of *nub*, *hth*, and *arau/caup*. In these cases, it seems likely that the transcription factors are integrated into different genetic regulatory modules, because these phenotypes are not similar to the phenotypes that result from disruption of body wall margin development.

**Presence of a wing serial homologue in the prothorax of Oncopeltus**

While the posterior and ventral supracoxal lobes, and the scutellar lobe, do not appear to be serially homologous with wings, we suggest that there is a portion of the prothoracic body wall sufficiently similar to wings to be called a wing serial homologue. This is the junction of the posterior pronotum margin with the posterior propleuron margin. It is located in the appropriate region of the body wall to be serially homologous with wings, arising at the most dorsal extent of the pleuron and the most distal extent of the notum. It was also affected more intensely by knockdown of *nub* and *vg* than other marginal tissues, causing the readily identifiable notched phenotype we observed in RNAi against these genes. But perhaps most importantly to the identification of serial homology, this particular portion of the prothorax is homeotically transformed into ectopic hemelytra by knockdown of the Hox gene *Sex-combs reduced* (Chesebro et al. 2009; Medved et al. 2015). While any one of these facts might not be sufficient to classify the pronotum/propleuron junction as a wing serial homologue, the confluence of them weighs heavily in favor of it. Indeed, the originators of the wing serial homologues concept have recently used similar evidence in describing a
variety of ectopic wing structures arising in *Cephalothorax* (the *Tribolium* ortholog of *Scr*) mutants (Clark-Hachtel et al. 2018).

*Model of wing origin and wing elaboration*

Our evidence prompts us to argue in favor of a model of wing origin and elaboration that posits a series of body plan innovations. In this model, the wing evolved in part through the action of a developmental module for flattened body wall evaginations. These would originally have been on all three thoracic segments, as suggested by fossil evidence (Kukalová-Peck 1983; Wootton and Kukalová-Peck 2000). The meso- and metathoracic lobes subsequently acquired novel identity as true wings by integration of tracheation and venation coupled with the acquisition of the direct and indirect flight muscles. As insect orders diversified, secondary co-option events led to elaborations such as the acquisition of hardened elytra of beetles by co-option of exoskeleton patterning (Tomoyasu et al. 2009), or the evolution of butterfly wing color forms by co-option of part of the eye patterning network (Martin et al. 2014). In other words, the body wall evagination developmental module produces novel body wall characters that become the substrate for subsequent phenotypic evolution.

In *Oncopeltus*, and perhaps Hemiptera broadly, this versatile body wall evagination module has been switched on during metamorphosis to create the dorsal and supracoxal lobes of the thoracic segments that we have discussed here. While the lobe-patterning module in *Oncopeltus* appears to be limited to the thorax, there is sufficient reason to doubt that all of the thoracic lobes are serially homologous with wings. However, the overall serial homology of the three thoracic segments contributes to developmental evolutionary constraints, or
correlated evolution (Liang et al. 2018). Because of this, the prothorax maintains tissue that is serially homologous to wings, the pronotum-propleuron junction.

**Acknowledgements**

Funding for this project was provided by grants from the University of Connecticut Department of Ecology and Evolutionary Biology to CRF, by a grant from the Evo-Devo-Eco Network (EDEN) to CRF, and by a grant from the National Science Foundation to ELJ.

**Intended publication and author contributions**

A manuscript based on this chapter is in preparation. Anticipated authors are C. R. Fisher, David R. Angelini, and Elizabeth L. Jockusch. CRF and ELJ conceived the experiments, CRF and DRA developed the protocol from DRA’s current laboratory techniques. CRF designed primers, cloned cDNA fragments, and constructed dsRNAs. DRA provided reagents. Injections were performed by CRF, with some assistance from Adam Chiu. Specimen preservation and scoring were performed by CRF. CRF drafted the manuscript and prepared the figures. ELJ and CRF edited the manuscript.

**Literature Cited**


Aspiras, A. C., F. W. Smith, and D. R. Angelini. 2011. Sex-specific gene interactions in the


Poelchau, M., C. Childers, G. Moore, V. Tsavatapalli, J. Evans, C. Lee, H. Lin, J. Lin, and K. Hackett. 2015. The i5k Workspace @ NAL — enabling genomic data access ,


Table 1 Initial primers generated for this study, for fragment amplification prior to cloning. Where three primers are listed for one gene, semi-nested PCR was used. The final column, f/r, indicates whether the primer is a forward primer (F) or reverse primer (R).

<table>
<thead>
<tr>
<th>name</th>
<th>gene</th>
<th>sequence (5’ – 3’)</th>
<th>f/r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Of’nub_30F</td>
<td>nub</td>
<td>GGCCGAAGATCTTACAAATGAC</td>
<td>F</td>
</tr>
<tr>
<td>Of’nub_3493R</td>
<td>nub</td>
<td>AGCATATCTGCCTCTTGACACACA</td>
<td>R</td>
</tr>
<tr>
<td>Of’tio_1018F</td>
<td>tio</td>
<td>ATGGTTAGGAAAGGTGCTGAA</td>
<td>F</td>
</tr>
<tr>
<td>Of’tio_2146R</td>
<td>tio</td>
<td>ACCTCTTCAGAAATGGAGCTG</td>
<td>R</td>
</tr>
<tr>
<td>Of’iro_F273</td>
<td>ara/caup</td>
<td>CTCACAGAGCAGAACCCCTAT</td>
<td>F</td>
</tr>
<tr>
<td>Of’iro_R764</td>
<td>ara/caup</td>
<td>ATCCTATGTCCTTTTCTCTG</td>
<td>R</td>
</tr>
<tr>
<td>Of’tailup_F63</td>
<td>tup</td>
<td>GCCAAACGAAAGATTTACCAC</td>
<td>F</td>
</tr>
<tr>
<td>Of’tailup_R866</td>
<td>tup</td>
<td>TCAGTAGAACGAGGGATGCACCA</td>
<td>R</td>
</tr>
<tr>
<td>Of’ap_F76</td>
<td>ap</td>
<td>TTATCGGCATAAAGAGGTGTG</td>
<td>F</td>
</tr>
<tr>
<td>Of’ap_R759</td>
<td>ap</td>
<td>ATCGGTTACCATCTGAAGCACT</td>
<td>R</td>
</tr>
<tr>
<td>Ofas'apR685</td>
<td>ap</td>
<td>GATGGTGTGGGTTGGGACG</td>
<td>R</td>
</tr>
<tr>
<td>Ofas'vgF1</td>
<td>vg</td>
<td>GATGTGCCTGTGACACAG</td>
<td>F</td>
</tr>
<tr>
<td>Ofas'vgR1</td>
<td>vg</td>
<td>AGAGATTCCGGGTTAGTAGGAC</td>
<td>R</td>
</tr>
<tr>
<td>Ofas'vgF2</td>
<td>vg</td>
<td>CTGTCTACACTGAGCGAGA</td>
<td>F</td>
</tr>
</tbody>
</table>
Table 2 T7-promoter tagged primers generated for this study. Lowercase nucleotides indicates sequence used for T7 promoter; uppercase indicates gene-specific sequence. The final column, f/r, indicates whether the primer is a forward primer (F) or reverse primer (R).

<table>
<thead>
<tr>
<th>name</th>
<th>gene</th>
<th>sequence (5’ – 3’)</th>
<th>f/r</th>
</tr>
</thead>
<tbody>
<tr>
<td>t7Oftio_2_F</td>
<td>tio</td>
<td>taatacgactcatactaggTGAGATCGATCACTGAAACTGG</td>
<td>F</td>
</tr>
<tr>
<td>t7Oftio_2_R</td>
<td>tio</td>
<td>taatacgactcatactaggCATAGCCTAATGTGATCCACGA</td>
<td>R</td>
</tr>
<tr>
<td>Of'T7Iro_F74</td>
<td>ara/caup</td>
<td>taatacgactcatactaggTGAAGGACAGCAGTGTACG</td>
<td>F</td>
</tr>
<tr>
<td>Of'T7Iro_R308</td>
<td>ara/caup</td>
<td>taatacgactcatactaggATCGCCAGCATGATCTTCTC</td>
<td>R</td>
</tr>
<tr>
<td>Ofas'T7ap_F1</td>
<td>ap</td>
<td>taatacgactcatactaggAAGATACCAGACGCTTTCTTC</td>
<td>F</td>
</tr>
<tr>
<td>Ofas'T7ap_R1</td>
<td>ap</td>
<td>taatacgactcatactaggCTGATGCTTGAAGGATGT</td>
<td>R</td>
</tr>
<tr>
<td>Ofas'T7vg_F1</td>
<td>vg</td>
<td>taatacgactcatactaggCTATCACCACCATCCCACTAC</td>
<td>F</td>
</tr>
<tr>
<td>Ofas'T7vg_R1</td>
<td>vg</td>
<td>taatacgactcatactaggGACCTCCACCCTACTCATATT</td>
<td>R</td>
</tr>
<tr>
<td>T7Ofas'tup_F1</td>
<td>tup I</td>
<td>taatacgactcatactaggGACCACGAAAGTTTCGAAAAAG</td>
<td>F</td>
</tr>
<tr>
<td>T7Ofas'tup_R2</td>
<td>tup I</td>
<td>taatacgactcatactaggAGTAGCAGGTCTCTCTTGATTG</td>
<td>R</td>
</tr>
<tr>
<td>Of'tup_T7F2</td>
<td>tup II</td>
<td>taatacgactcatactaggGTGACATCTTATCGAGCCG</td>
<td>F</td>
</tr>
<tr>
<td>Of'tup_T7R2</td>
<td>tup II</td>
<td>taatacgactcatactaggTATCAGCTGAGGTTGGAGCA</td>
<td>R</td>
</tr>
</tbody>
</table>
**Table 3** Experimental and control group sizes and mortality for RNAi experiments. Instars indicates the number of instars at which each individual was injected; \(N_{\text{inj}}\) and \(N_{\text{adult}}\) are the number of individuals injected and surviving to adulthood, respectively. For the gene *tup*, two different non-overlapping fragments of the full-length transcript were used to verify that results were not due to off-target effects, and these are referred to as *tup*-I and *tup*-II.

<table>
<thead>
<tr>
<th>gene</th>
<th>instars</th>
<th>(N_{\text{inj}})</th>
<th>(N_{\text{adult}})</th>
<th>mortality</th>
<th>(N_{\text{inj}})</th>
<th>(N_{\text{adult}})</th>
<th>mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ap</em></td>
<td>2</td>
<td>22</td>
<td>14</td>
<td>35%</td>
<td>22</td>
<td>17</td>
<td>23%</td>
</tr>
<tr>
<td><em>ara/caup</em></td>
<td>1</td>
<td>10</td>
<td>8</td>
<td>20%</td>
<td>10</td>
<td>9</td>
<td>10%</td>
</tr>
<tr>
<td><em>ara/caup</em></td>
<td>2</td>
<td>60</td>
<td>7</td>
<td>88%</td>
<td>29</td>
<td>24</td>
<td>17%</td>
</tr>
<tr>
<td><em>hth</em></td>
<td>1</td>
<td>73</td>
<td>33</td>
<td>55%</td>
<td>29</td>
<td>26</td>
<td>10%</td>
</tr>
<tr>
<td><em>nub</em></td>
<td>1</td>
<td>40</td>
<td>17</td>
<td>58%</td>
<td>25</td>
<td>20</td>
<td>20%</td>
</tr>
<tr>
<td><em>tup</em>-I</td>
<td>1</td>
<td>13</td>
<td>10</td>
<td>23%</td>
<td>not done</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>tup</em>-I</td>
<td>2</td>
<td>9</td>
<td>5</td>
<td>44%</td>
<td>not done</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>tup</em>-II</td>
<td>1</td>
<td>10</td>
<td>9</td>
<td>10%</td>
<td>10</td>
<td>4</td>
<td>60%</td>
</tr>
<tr>
<td><em>tio</em>-II</td>
<td>2</td>
<td>20</td>
<td>15</td>
<td>25%</td>
<td>13</td>
<td>12</td>
<td>8%</td>
</tr>
<tr>
<td><em>vg</em></td>
<td>2</td>
<td>20</td>
<td>4</td>
<td>80%</td>
<td>10</td>
<td>9</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>13</td>
<td>13%</td>
<td>11</td>
<td>11</td>
<td>0%</td>
</tr>
</tbody>
</table>
Table 4 Penetrance for major RNAi-induced body wall and wing phenotypes. For *ara/caup* and *tup*, penetrance has been calculated based on experimental individuals surviving to adulthood. The sample size is shown below each gene. “NS” means that the phenotype was not scored.

<table>
<thead>
<tr>
<th>phenotype</th>
<th>ap</th>
<th><em>ara/caup</em></th>
<th>hth</th>
<th><em>nub</em></th>
<th><em>tio</em></th>
<th><em>tup</em></th>
<th>vg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(14)</td>
<td>(15)</td>
<td>(33)</td>
<td>(17)</td>
<td>(4)</td>
<td>(24)</td>
<td>(13)</td>
</tr>
<tr>
<td>Pronotum size change</td>
<td>57%</td>
<td>26%</td>
<td>45%</td>
<td>75%</td>
<td>75%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Collar reduced/retracted</td>
<td>57%</td>
<td>13%</td>
<td>45%</td>
<td>25%</td>
<td>75%</td>
<td>0%</td>
<td>15%</td>
</tr>
<tr>
<td>Pronotum midline reduced</td>
<td>57%</td>
<td>27%</td>
<td>45%</td>
<td>NS</td>
<td>75%</td>
<td>63%</td>
<td>NS</td>
</tr>
<tr>
<td>Pronotum corners reduced</td>
<td>57%</td>
<td>27%</td>
<td>0%</td>
<td>NS</td>
<td>NS</td>
<td>52%</td>
<td>NS</td>
</tr>
<tr>
<td>Pronotum-pleural junction reduced</td>
<td>36%</td>
<td>7%</td>
<td>18%</td>
<td>75%</td>
<td>NS</td>
<td>0%</td>
<td>77%</td>
</tr>
<tr>
<td>Reduction of T1 pleural margin</td>
<td>7%</td>
<td>14%</td>
<td>24%</td>
<td>71%</td>
<td>75%</td>
<td>0%</td>
<td>46%</td>
</tr>
<tr>
<td>Reduction of T1 pleural lobes</td>
<td>7%</td>
<td>0%</td>
<td>24%</td>
<td>75%</td>
<td>75%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Mesoscutellum shape change</td>
<td>43%</td>
<td>7%</td>
<td>36%</td>
<td>6%</td>
<td>75%</td>
<td>79%</td>
<td>100%</td>
</tr>
<tr>
<td>Reduction of T2 pleural margin</td>
<td>0%</td>
<td>14%</td>
<td>24%</td>
<td>19%</td>
<td>75%</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td>Reduction of T2 pleural lobes</td>
<td>0%</td>
<td>0%</td>
<td>24%</td>
<td>63%</td>
<td>75%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Reduction of T3 pleural margin</td>
<td>0%</td>
<td>14%</td>
<td>24%</td>
<td>19%</td>
<td>75%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Reduced/defective scent groove</td>
<td>0%</td>
<td>0%</td>
<td>15%</td>
<td>0%</td>
<td>NS</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Smaller wings</td>
<td>0%</td>
<td>20%</td>
<td>0%</td>
<td>75%</td>
<td>NS</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Desclerotized hemelytra</td>
<td>86%</td>
<td>NS</td>
<td>0%</td>
<td>0%</td>
<td>NS</td>
<td>21%</td>
<td>0%</td>
</tr>
<tr>
<td>Thin/depigmented hind wings</td>
<td>50%</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
<td>21%</td>
<td>0%</td>
</tr>
<tr>
<td>Defective wing hinges</td>
<td>NS</td>
<td>NS</td>
<td>15%</td>
<td>0%</td>
<td>75%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Figure 1. RNAi effects on posterior and dorsal pleural margins. A. negative control pterothorax with prothorax removed, showing normal scent groove, wing-support processes, and posterior reduplications (lobes). B. RNAi against *Ofas’hth* causes extreme reduction of pleural body wall margins. Arrowhead points to exposed T3 spiracle. Asterisk indicates highly reduced external scent groove. C. RNAi against *Ofas’nub* causes similar but less extreme pleural margin reduction, but does not affect the external scent groove. ep2, ep3 = epimeral protrusion of T2 and T3 (also called notal processes); Rd2, Rd3 = posterior reduplication of T2 and T3; sg = external scent groove; sp3 = spiracle of T3; wp2 = pleural wing process of T2.
Figure 2. Effects of RNA interference on supracoxal lobes.
A. Wild type supracoxal lobes of mesothorax. Lobes meet along nearly their entire length and cover the trochantin and proximal coxa. B. RNAi against Ofas’nub causes an open coxal cleft due to the reduction of supracoxal lobes. C, D. RNAi against Ofas’hth and Ofas’pio result in more extreme reductions of the supracoxal lobes. Reduction is so severe in Ofas’pio RNAi specimens that the trochantin (tr) is visible.
Figure 3 RNAi interference effects on the anterior prothoracic margin ("collar"). A. Negative control specimen showing wild type. The collar is a folded lobe of the prothoracic body wall (notum, pleuron, and sternum) and normally extends to just behind the eyes. B. Specimen with Of’’thr knockdown phenotype. Collar is retracted and collum is exposed. C. Knockdown of Of’’ap causes the dorsal but not ventral portion of the collar to become retracted. D. Semi-lateral view of same specimen, showing retracted dorsal portion of collar. Arrowheads point to anterior margin of prothorax. Double-stroked cross in B. marks the enlarged mandibular plate that occurs from Of’’thr knockdown. Asterisk in D. shows ventral/pleural extent of collar unaffected by knockdown of Of’’ap. clm = collum
Figure 4 Dorsal view of pronotum and scutellum under every treatment in this study. pn = pronotum; sctl = mesoscutellum; asterisk denotes reduction or aberration of pronotum midline; white arrowheads denote missing or reduced posterior pronotum corners; arrows denote expanded or shape-changed scutellum; dashed line on C. shows scutellum reduction that occurs with knockdown of Ofas’tup.
Figure 5. Differential reduction of the pronotum-propleuron junction under knockdown of Ofas'nub and Ofas'vg. Dotted line follows the most posterior edge of the prothoracic body wall margin. B. RNAi targeting Ofas'nub results in an overall smaller prothoracic evagination, but reduces the propleuron and the junction of the propleuron more severely, causing a characteristically extreme curve when viewed laterally. C. RNAi against Ofas'vg resulted in a similarly overall smaller prothorax phenotype, but the posterior margin of the propleuron is not reduced as severely as the junction of the pronotum and propleuron, resulting in a notched lateral prothoracic margin.
Figure 6. Depletion of *Ofas’ap* results in shape changes in the mesoscutellum due to a reduction of the dorsal but not ventral layer of the lobe. **A.** Semi-lateral view specimen with RNAi targeting *Ofas’ap*. The scutellum points up and away from the body instead of being domed and flush with the wings. **B.** Full dorsal view of another *Ofas’ap* RNAi specimen. Dashed line indicates the detail zoomed in **C.** In **C**, the arrowhead points to the cleft that forms at the tip of the mesoscutellum as a result of the dorsal layer being reduced relative to the ventral layer and pulling the edges away from each other. **D.** shows a wild type mesoscutellum at the same magnification. Note the loss of sensory setae in the *Ofas’ap* RNAi phenotype in **C.**
**Figure 7.** Effects of RNA interference on wing size, shape, and sclerotization. A. wild-type forewing. cl=clavus, cor=corium, mem=membranous portion of the forewing. B. Knockdown of *Ofas’ap* results in a desclerotized wing with a clear membranous window in the center (marked by the white dashed line.) C. Forewing of *Ofas’ara/caup* RNAi specimen. Overall length is shorter and distal wing blade is disturbed. D. *Ofas’nub* RNAi forewings are extremely short, with reduction of corium and clavus being more severe than the reduction of the membranous portion. Wing venation is generally undisturbed in membranous portion, but appears truncated. E. Knockdown of *Ofas’tup* does not affect wing size or shape, but does result in small depigmented areas of the forewing, particularly in intervein regions. F. *Ofas’vg* RNAi results in much smaller forewings, with the most distal part of the membranous wing most affected.
Figure 8. RNAi targeting *Ofas'vg* affects distal antennae in addition to wings and body wall margins. Fl.3=third flagellomere, Fl.2=second flagellomere. The dashed box included to show that Fl.2 in negative control and *Ofas'vg* RNAi specimens are scaled to the same proportion. B. In 100% of specimens in the *Ofas'vg* RNAi experiment, the most distal flagellomere was small, distorted, and lacking setae.
Chapter 4: Development of functional genomic analysis tools in an emerging model

treehopper, *Entylia carinata*

Abstract

Testing hypotheses of the origin of novel morphologies such as the treehopper helmet requires functional, interventionist experiments to understand the role of genes during development. RNA interference (RNAi) is a powerful method of decreasing expression of target genes, allowing for reverse genetics without the development of transgenic lines. RNAi is mediated through cellular machinery that is conserved across Eukaryota, but recent efforts in diverse insect taxa demonstrate that there are some taxon-specific differences that affect the robustness of the RNAi response. In this chapter, we report for the first time the successful knockdown of target genes by RNAi in a treehopper, *Entylia carinata*. We demonstrate that *E. carinata* expresses all of the requisite genes for the proper functioning of the RNAi pathway, present a detailed injection protocol, and show that RNAi is accomplished for genes in the pigment pathway. However, attempts to produce knockdown phenotypes for developmental genes were unsuccessful. Based on recent literature assessing the spotty success of RNAi in various insect taxa, we provide a path forward for optimization of RNAi in *E. carinata*.

Introduction

The treehopper helmet is an evolutionary novelty in that it has diverged morphologically from its precursor and assumed a new functional role. As defined by Wagner (2015), this puts it in the category of a Type II novelty, because it is an existing body part that has undergone drastic physical transformation, rather than a Type I novelty, which is
a wholly new body part with no easily discernable precursor. In other words, the treehopper helmet, while clearly very different from the homologous flat pronotum in other insects, retains its identity as the pronotum.

The literature on novelty suggests three hypotheses concerning changes in the regulation of genes that pattern the first thoracic body wall and these hypotheses are discussed in detail in the first chapter of this dissertation. Our analysis of gene expression by comparative RNA sequencing (Chapter 1) strongly supports the wing co-option hypothesis. However, while our RNA-seq expression data conform with the hypothesis of wing patterning co-option, there is an alternative (though not necessarily mutually exclusive) hypothesis that cannot be supported or rejected with gene expression data alone, and that is that the treehopper helmet is a partial wing serial homologue (Tomoyasu et al. 2017).

This hypothesis builds on the insect wing dual origin hypothesis, an emerging view of wing evolution in which pleural body wall tissue combined with tergal body wall tissue to create the wings (Niwa et al. 2010; Clark-Hachtel et al. 2013; Ohde et al. 2013; Medved et al. 2015; Elias-Neto and Belles 2016). While the dual origin hypothesis has historical antecedents, it has come into prominence in the past decade because of an increased attention paid to the development of wings in insect taxa outside of Drosophila, including coleopterans (Clark-Hachtel et al. 2013; Ohde et al. 2013), blattodeans (Elias-Neto and Belles 2016), and hemipterans (Medved et al. 2015). Three other, general hypotheses for wing evolution are the gill hypothesis (Wigglesworth 1973; Averof and Cohen 1997; Wootton and Kukalová-Peck 2000), the exite (leg branch) hypothesis (Kukalová-Peck 1978, 1983; Wootton and Kukalová-Peck 2000), and the paranotal lobe hypothesis (Crampton 1916; Snodgrass 1927; Matsuda 1970), each of which holds that wings evolved from the eponymous organ. The dual
origin hypothesis holds that wings, a Type I novelty, formed as a kind of hybrid of two different tissues. It further implies that in non-winged segments of extant insects, the ancestral tissue-types are presumably still present, and are partially, rather than wholly, serially homologous to the wings (Clark-Hachtel et al. 2013; Ohde et al. 2013; Tomoyasu et al. 2017).

Under the dual origin hypothesis, the treehopper helmet could be a partial wing serial homologue in virtue of its identity as prothoracic tergum (Stegmann 1998; Mikó et al. 2012). And, given this possibility, it may be difficult to ascertain from differential gene expression if the similarity between the helmet and wings exists because of co-option of wing-patterning genes, or because wing-patterning genes were ancestrally expressed at a small portion of the prothorax that has undergone differential growth in the evolution of the helmet. If the insect wing dual origin hypothesis is correct, then, given our other evidence, one of three possible scenarios could be true for the origin of the treehopper helmet: 1) The helmet arose by wing co-option but is not a wing serial homologue; 2) the helmet is a wing serial homologue, because a portion of the insect pronotum is a wing serial homologue, and modulation of genes that were ancestrally involved in patterning both this portion of the pronotum and wings led to the helmet’s evolution, or 3) the a portion of the insect pronotum is a wing serial homologue, and also a secondary co-option of the wing patterning network was required to give rise to the novel treehopper helmet.

Distinguishing between these three scenarios requires comparative functional evidence about the developmental role of the genes that are differentially expressed in tandem in the treehopper helmet and body wall. To date, no methods have been reported for functional developmental genetics in any treehopper species. Therefore, we present here a
method for achieving gene silencing through RNA interference in the widespread, common, and lab-friendly camelback treehopper, *Entylia carinata* (Forster 1771). We chose RNA interference as our reverse genetics method because it is not transgenic (Fire et al. 1998; Mello and Conte 2004), and so it can be implemented at the exact developmental stage of interest, allowing embryonic and early instar nymphs to develop normally until the later nymphal stages, in which adult morphology is attained (Mito et al. 2010). Further, gene silencing by RNA interference relies on a highly conserved (pan-eukaryote) cellular pathway, and has been successful not only in a wide variety of insects (Dong and Friedrich 2005; Howard et al. 2006; Liu et al. 2010), but also in a hemipteran closely related to treehoppers, the leafhopper *Homalodisca vitripennis* (Rosa et al. 2010, 2012). Because we have developed transcriptomic resources for *E. carinata*, and because this species is a multivoltine, year-round breeder with a relatively fast life history and is easily reared on sunflowers in a greenhouse, we believe that the methods presented here build the basis for a powerful tool to unlock the mysteries of the origin of the novel treehopper helmet.

**Methods**

We chose to target the pigment pathway for our methods development because the predicted phenotypes are simple and unambiguous. The insect pigment pathway is well studied and appears to be highly conserved (True 2003; Futahashi et al. 2005; Shirataki et al. 2010; Ferguson et al. 2011; Liu et al. 2014). To knock down melanin production, we chose to target the gene *tyrosine hydroxylase* (*TH*), which codes for an enzyme that converts tyrosine into a dopamine precursor, a rate-limiting step of the melanin synthesis pathway. To increase melanin production, we chose the gene *ebony* (*e*) which codes for an enzyme that converts
dopamine and beta-alanine to a yellow pigment (Ferguson et al. 2011). Depletion of this enzyme redirects the pigment precursor to the eumelanin pathway, producing the eponymous ebony phenotype.

To test our method on developmental processes, we chose four genes: Sex-combs reduced (Scr), a Hox gene that controls patterning in the first thoracic segment (Pattatucci et al. 1991; Rogers et al. 1997) and for which knockdown phenotypes are highly predictable; apterous (ap), a gene in the canonical wing patterning pathway (Cohen et al. 1992; O’Keefe and Thomas 2001; Prakash and Monteiro 2018) that we have investigated in a related hemipteran, Oncopeltus fasciatus, (Chapter 3) and which our RNA-Seq data show are upregulated in Entylia helmets and wings (Chapter 1); Distal-less (Dll), a gene that in Drosophila and other arthropods is responsible for establishing the distal portion of the proximal-distal axis of appendages such as legs (Vachon et al. 1992; Moczek et al. 2006; Sharma et al. 2013), and thus is a good candidate for testing the leg co-option hypothesis for the helmet; and homothorax (hth), a Hox co-factor known to be involved in the development of body wall characters (Casares and Mann 2000; Smith and Jockusch 2014), which we have also investigated in O. fasciatus.

Testing for presence of RNAi machinery in Entylia

Our protocol for RNAi involves transcribing double-stranded RNA (dsRNA) from a cloned gene of interest and injecting it into the hemolymph of an insect. Once injected, the dsRNA is taken up by cells and prevents production of the gene’s protein product (Fire et al. 1998; Scott et al. 2013; Wilson and Doudna 2013). When RNAi is used to temporarily silence genes during an organism’s development, phenotypes may result from which the silenced gene’s function may be inferred.
Verifying that *E. carinata* both possesses and expresses the necessary genes for RNA interference involves identifying likely transcripts by similarity using BLAST (Altschul et al. 1990), checking the predicted peptides of those transcripts for the necessary conserved domains, and constructing gene trees to support orthology relationships. We downloaded the amino acid sequences for *Drosophila melanogaster* genes *dicer-1, dicer-2, dicer-3, drosha, r2d2, loquacious, argonaute-1, argonaute-2, argonaute-3*, and *aubergine* from UniProtKB. These sequences were used as queries in BLAST searches against our custom annotated transcriptome for *E. carinata* (Chapter 1). The top hit (or the top three, if their scores were similar) were used as queries in BLAST searches against the nr database (NCBI) limited to *Drosophila melanogaster* proteins, and only reciprocal best BLAST hits were used in further analyses. For these transcripts, we retrieved the matching peptide from the annotated proteome that we had previously constructed. We uploaded these amino acid sequences to ScanProsite (Castro et al. 2006) to identify conserved domains. We created alignments using the MUSCLE algorithm implemented in MEGA7 (Kumar et al. 2016) for the *E. carinata* candidate sequences with the putative orthologous/paralogous proteins from *Drosophila, Tribolium castaneum*, and *Nilaparvata lugens* (the brown rice planthopper, a hemipteran related to *E. carinata*). We trimmed the amino acid alignments to the conserved domains identified by ScanProsite, and constructed neighbor-joining trees using the Dayhoff (1978) model of amino acid substitution.

**RNA interference candidate gene selection, amplification, and sequencing**

To develop our injection protocol, we started with the gene sequences from *Drosophila* for *tyrosine hydroxylase* and *ebony*, two genes in the insect melanin synthesis pathway (Fig. 7). Using a similar to that described above for identifying core RNAi
machinery genes, we found the orthologous transcripts in our *Entylia* transcriptome. Gene-specific primers were designed for each gene (Table 1). Amplicons were used to transform *E. coli* and plasmid DNA was extracted from resulting clones. We sequenced the inserts from the plasmid DNA to verify successful transformation with the target sequence fragment, and then used the exact sequence to design primers with T7 promoter sequences. The T7 promoter primers were used to amplify template DNA for transcription to dsRNA using the HiScribe kit (NEB). Purified dsRNA was resuspended in 30 µL nuclease free water, quantified via Nanodrop spectrometry, and electrophoresed through a 1% agarose non-denaturing gel to verify product integrity and size.

dsRNA design for developmental genes

Because *E. carinata* is not a developmental model organism, there were no genomic or transcriptomic resources available for it prior to the transcriptome produced by our studies (Chapter 1). For this reason, the genes Ec’ap, Ec’Scr, and Ec’Dll were initially amplified by two rounds of degenerate PCR. For Ec’ap and Ec’Scr, degenerate primers were designed using the CODEHOP web tool (now implemented as part of the software Base-to-Base version 3) (Rose et al. 2003; Boyce et al. 2009; Tu et al. 2018). To design degenerate primers with CODEHOP, we aligned the amino acid sequences for orthologs in other insect species including hemipterans and uploaded this alignment to the web tool. We selected primers from the suggested pool to balance degeneracy against sufficient length for dsRNA construction of the predicted amplicons (Table 1). For Ec’Dll, we used a degenerate universal arthropod primer (a gift from Frank Smith). Initial PCR and cloning of Ec’hth was begun after a reference transcriptome was available, and gene-specific primer design was performed as described above for pigment pathway genes.
PCR amplicons were inserted into plasmid vectors and used to transform *E. coli* whose plasmid DNA were sequenced as described above. Exact sequences were used to design primers with T7 promoters for dsRNA transcription (Table 2).

**dsRNA injection protocol**

Previous studies in Hemiptera have identified numerous difficulties in RNA interference (Li et al. 2013; Scott et al. 2013; Christiaens et al. 2014), most of which amount to either no gene silencing at all, or gene silencing with a milder and shorter-duration response relative to the robust and long-lived response found in some other insects, such as *Tribolium* (Christiaens et al. 2014). Researchers have achieved successful RNAi in *O. fasciatus* and *Homalodisca vitripennis* (a related hemipteran) using high concentrations of dsRNA (between 1.2 to 4 µg/µL) and two rounds of injections (Chesebro et al. 2009; Rosa et al. 2012; Medved et al. 2015). We applied this approach to *E. carinata*. For *Ec’Th* dsRNA and *Ec’e* dsRNA, we used a concentration of 5.5 µg/µL in physiological buffer (0.01 mM NaPO₄, 5 mM KCl, and 0.05% McCormick green or blue food coloring; red food coloring (McCormick) was also explored, but proved lethal.) Due to the size and morphological changes that naturally occur between 4th and 5th instar treehoppers, we used different injection sites for the two injections. Fourth instar treehoppers were injected between the third thoracic segment and first abdominal segment—the easiest soft part of the body to access in that stage. In the 5th instar, the developing helmet bud overhangs the abdominal segments and makes this site less accessible, but the nymphs are large enough to tolerate injection in the membrane of the coxal cavity of the right first thoracic leg. Pulled glass capillary needles, mounted on a micro-manipulator and driven by a manually controlled syringe, were used for both stages. Negative controls were injected at the same time as
experimental groups. Injected nymphs were maintained in organza draw-string bags around a
growing sunflower plant leaf in the EEB greenhouse.

Verifying RNAi success via qualitative PCR

In the case of RNAi against \( TH \), we made cDNA for three experimental individuals
and three negative control individuals. cDNA reactions were performed with the Qscript Flex
kit (Quanta) using oligo(d)T primers to preferentially reverse-transcribe poly-adenylated
messenger RNA transcripts. We used these cDNAs as templates for PCR for the target gene,
\( TH \), and for a control gene not targeted in the \( TH \) experiment (Ec 'Scr).

Statistical support for RNAi phenotypes and lethality

A logistic regression was performed with each paired set of control and experimental
data using the lm() function in R (RStudio Team 2015; R Development Core Team 2016).
This was performed on the mortality data for all genes, and on the phenotype data for
pigment pathway genes.

Specimen preservation and phenotype scoring

For specimens from the pigment pathway experiments, scoring was performed on
individuals after they had been allowed one to three days to fully harden and develop
pigments. Instead of being stored in 70% ethanol, which might affect pigmentation, these
individuals were placed in empty, labeled Eppendorf tubes and frozen until they could be
scored. Pigmentation was assessed on frozen individuals. Scored specimens were preserved
as dry collection specimens, pointed and pinned, labeled with date, experiment number and
specimen id, and treatment.

For the developmental gene experiments, we developed scoring matrices for external
characters based on \( Entyilia carinata \) wild-type morphology. We scored adult specimens from
each experimental cohort alongside the corresponding negative control cohort, assigning 0 for normal morphology, and 1 for aberrant morphology. These specimens were preserved in 70% ethanol.

In a few rare cases, nymphs from experiments seemed to present evidence of knockdown prior to eclosion as adult. In these instances, photographs and careful notes were taken of the evidence before preserving these nymphs in RNALater.

**Results**

*Entylia carinata* expresses genes for insect core RNAi machinery

We identified orthologues in *E. carinata* for *Dicer-2*, *Argonaute 2*, and *r2d2*, three genes whose products are required for the intracellular components of short-interfering RNA (siRNA) induced silencing (He and Hannon 2004; Mello and Conte 2004; Tomoyasu et al. 2008; Bartel 2009; Wilson and Doudna 2013). Additionally, we identified orthologues of *Dicer-1*, *Dicer-3*, *Argonaute 1*, *Argonaute 3*, *aubergine*, and *loquacious*, genes whose protein products are typically involved in transcriptional silencing through the micro RNA (miRNA) or piwi (piRNA) pathways (Tomoyasu et al. 2008; Ghildiyal and Zamore 2009).

The predicted proteins for the *E. carinata* RNAi machinery orthologues—*Ec’Dcr-2*, *Ec’Ago 2*, and *Ec’r2d2*—contain the necessary domains for processing exogenous dsRNA, converting it to small interfering RNA (siRNA), and loading it into the RNA-induced silencing complex (RISC) (*Fig 2, 3*). Additionally, gene trees constructed from the predicted amino acid sequences of these three genes and other genes in their families support the identity established by the domain architecture (*Fig. 4-6*).
Survivability of injection protocol

Our mean survival rate post injection was 71% (sd 16%) for injections at L4 and 75% (sd 17%) for injections at L5, leading to an overall survival rate of 53% (18%) for our consecutive injection protocol.

**RNA interference against pigment pathway genes indicates successful knockdown of transcripts**

_E. carinata_ are known to display a variety of helmet color morphs within a population (Matausch 1910), ranging from mostly brown to a mottled tan, therefore helmet color morphs were scored as either pale or dark. The occurrence of pale helmet color morphs was significantly different in _Ec' TH_ injections compared to control injections (p=0.003, Table 5). Leg coloration is not as variable as helmet coloration, with wild-type adults possessing tan-colored legs with dark brown cuculate setae on the metathoracic tibia (Fig. 1B); in individuals displaying a phenotype for RNAi against _Ec' TH_, cuculate setae were much paler (Fig. 1D). Semi-quantitative PCR of cDNA from experimental and control-injected individuals verified that the target mRNA was knocked down (Fig 8). Lethality was not different between negative controls and experimental cohorts (Table 3). Results for the experiment targeting _Ec’ e_ showed relatively low penetrance (30%, p=0.0569 n.s., Table 4), but the phenotype was unmistakable: black instead of brown, due to the lack of yellow or tan pigment (Fig. 1E-F). Regardless of color morph, _E. carinata_ helmets range in color from light tan to dark brown, but never completely black (Fig. 1A). Further, wild-type adult legs are always light tan and never black (Fig. 1B). In our phenotypic _Ec’e_ RNAi, the legs were black, and the dark spot on the wings (normally a dark brown) was black and larger than
typical. These results are similar to those reported for dsRNA-induced knockdown of *Tc’e* in *Tribolium castaneum* (http://ibeetle-base.uni-goettingen.de/details/TC011976).

RNA interference against developmental genes was largely unsuccessful

With a few notable exceptions, experiments attempting to generate developmental phenocopies from the knock down of *Sex-combs reduced, Distal-less, apterous*, and *hth* were unsuccessful (Table 3).

*Increased mortality with RNAi against Ec’Scr and Ec’Dll*

For experimental cohorts for the genes *Ec’Scr* and *Ec’Dll*, mortality was significantly increased compared to control. *Ec’Scr* RNAi resulted in 82% mortality (= percent of injected 4th instar nymphs that did not survive to adulthood) compared to 50% mortality in the negative control (p=0.0325, Table 3). Injection with dsRNA targeting *Ec’Dll* resulted in 65% mortality compared to 42% mortality in the negative control group (p=0.0493, Table 3).

*Developmental arrest at low penetrance for RNAi against Ec’Scr*

In the *Ec’Scr* experiment, 3 of the 28 nymphs injected at the L4 stage appeared to molt into a novel instar retaining L4 morphology but with a size intermediate between L4 and L5, a phenotype we denote as L4’. Two of these were placed in RNALater when discovered in this state; the remaining individual so affected was left to continue development on a sunflower cutting. However, it never molted again, neither into an adult nor into a true L5, and was preserved in ethanol when it died 27 days post initial injection. The normal stadium length for *Entylia carinata* 4th and 5th instar is 5-7 days, so this individual represented an extreme developmental outlier. Developmental arrest was never observed in any negative control, and nor was the L4’ phenotype. Similar results have been reported in RNA interference studies with nymphs of the hemipteran species *O. fasciatus*.

164
(Erezyilmaz et al. 2006) and *Pyrrhocoris apterus* (Konopova et al. 2011), when injected at the 4<sup>th</sup> instar with dsRNA targeting the Broad-Complex (*BR-C*) genes, which are involved in mediating the function of juvenile hormone during insect metamorphosis. We have also observed developmental arrest (but not juvenilized instars) in knockdown of *nubbin* in *O. fasciatus* (unpublished data).

**Low penetrance with patchy effects in RNAi targeting Ec’Dll**

A small number of adults (n=2 of 19) in the experimental cohort for RNAi against Ec’Dll presented with leg phenotypes consistent with the predicted role of *Distal-less* in leg development. These phenotypes included one adult with drastic shortening of one leg, resulting from fusion and reduction of leg segments; and one adult with a leg missing tarsomeres, in which pretarsal claws were attached directly to the end of the tibia (Fig. 9B).

**Discussion**

*RNA interference is a viable method for functional genomic analysis in Entylia carinata*

The results of our investigation demonstrate that *E. carinata* possesses all the necessary cellular machinery needed for RNA interference, that introduction of dsRNA into the hemolymph by injection is well tolerated by both 4<sup>th</sup> and 5<sup>th</sup> instar nymphs, and that dsRNA so introduced is capable of inducing mRNA depletion. Results for RNAi targeting Ec’TH were less striking than expected compared to similar studies (Liu et al. 2014), but were nonetheless convincing due to clearly diminished amounts of dark pigments in the legs of phenotypic specimens (Fig. 1C-D). We observed no clear phenotypic effects from developmental gene experiments aside from increased mortality in RNAi targeting Ec’Scr and Ec’Dll (compare, for example, our previously reported success in nymphal RNAi with *O. fasciatus* with two of the genes investigated here, *ap* and *hth*). This indicates that additional
optimization is needed before RNAi can be employed to gather the functional data necessary to test our hypotheses.

Important considerations for future optimization include ascertaining why pigment pathway RNAi was clearly successful, when developmental gene RNAi was not. Unfortunately, this may not be an easy task. Differential phenotype penetrance as a function of target gene is nothing new in insect RNAi (Angelini et al. 2005; Schmitt-Engel et al. 2015). Even in “well-behaved” model systems such as Tribolium, semi-candidate gene screening approaches as well as genome-wide screens have yielded unexpectedly low penetrance or complete absence of anomalous phenotype for some genes and unexpectedly severe phenotypes for others (Linz and Tomoyasu 2015).

These differential success rates may be influenced by a variety of biological factors, including failure due to low protein turnover or low activity thresholds. For example, Rinkevich and Scott (2013) reported that dsRNA injections of Tribolium pupae targeting the gene coding for the alpha-6 nicotinic acetylcholine receptor (α6) resulted in quantifiable reduction of the targeted mRNA, but did not result in the expected pesticide-resistance phenotype, likely because the α6 protein is highly stable or not required in high numbers to mediate pesticide susceptibility. A similar set of circumstances may explain why our pigment pathway RNAi succeeded where development gene RNAi failed. The synthesis of N-β-alanyldopamine (NBAD, the precursor to yellow pigments) from hemolymph-circulating dopamine occurs rapidly in the insect cuticle just after molting, and requires the enzyme product of the gene ebony. In caterpillars of the swallowtail butterfly Papilio xuthus, expression of the gene ebony is undetectable until the 12th hour after molt, and then only expressed in a highly restricted domain coinciding with the small reddish-brown pigmented
area of the caterpillar (Futahashi et al. 2005), implying that its protein product is both short-lived and produced on a just-in-time basis. Hox genes such as Sex-combs reduced are expressed throughout an organism’s development because they are part of the underlying positional information that determines spatial expression domains (Wagner 2014). Therefore, the choice of target gene should be made carefully for future efforts, and to the extent possible, temporal-spatial expression and threshold functional levels should be taken into account.

Other factors that may affect the differential success of RNA interference include the expression level of the target, the length of the dsRNAs employed, the specific developmental window during which RNAi is employed, and a variety of as-yet unknown specific details about the RNA interference mechanisms in E. carinata (Mugat et al. 2003; Terenius et al. 2011; Li et al. 2013; Scott et al. 2013; Wang et al. 2016). The next section addresses some possible reasons for the lack of developmental phenotypes, and suggests avenues for testing and overcoming these biological difficulties.

**Options for future optimization of RNA interference in Entyilia carinata**

Because of the specificity and low toxicity for non-target species, RNA interference is of increasing interest to pest management research (Li et al. 2013; Scott et al. 2013; Christiaens et al. 2014; Hajeri et al. 2014; Wynant et al. 2014; Ramesh Kumar et al. 2016). However, many of the species of most concern have proven stubbornly resistant to gene expression knockdown by RNAi, for example, the yellow fever vector Aedes aegypti, the ubiquitous garden pests of Aphis spp., the Chagas disease vector Rhodnius prolixus, and most members of Lepidoptera (Terenius et al. 2011; Mysore et al. 2013; Christiaens et al. 2014). There are three primary suspects for the cause of RNAi insensitivity in these insects: high

The siRNA interference pathway is hypothesized to have evolved as an innate cell-immunity response to viral infection (Karlikow et al. 2014). Therefore, high levels of viral infection may leave insects unable to mount a response to investigatory injections of dsRNA, simply because the cellular machinery is already occupied (Smagghe et al. 2014; Swevers et al. 2016; Santos et al. 2018). This has recently been demonstrated to be the cause behind RNAi resistance in cell lines of lepidopterans (Swevers et al. 2016). This phenomenon would explain why RNA interference works so well in *O. fasciatus*, a seed predator, which would be expected to come into contact with fewer viruses, and in *Tribolium castaneum*, the flour beetle, which feeds on dried plant material rather than live. Determining if viral load RNAi inhibition is the case in *E. carinata* is beyond the scope of the current study, but methods of clearing viruses from the lab colony and preventing further viral infection could be explored.

The role of exo- and endonucleases in preventing RNAi responses has been examined in a number of insects. In aphids (Christiaens et al. 2014), silkworms (Garbutt et al. 2013), and locusts (Wynant et al. 2014), dsRNA is rapidly degraded upon exposure to cell-free hemolymph and gut extracts. Similarly, degradation of dsRNA by cell-free insect hemolymph was demonstrated in a number of insects across insect orders and appears to be a highly variable trait hemolymph extract from *Tribolium* did not degrade dsRNA even at relatively high concentration (16 mg/mL), while hemolymph extract from other beetles including *Popillia japonica* and *Coccinella septempunctata* degraded dsRNA at relatively low concentrations (0.125 mg/mL and 1 mg/mL fluid, respectively) (Singh et al. 2017). Based on the evidence that RNA interference in *E. carinata* seems to require very high
dosage and double injections, it is possible that nucleases are to blame for the lack of vigorous RNAi response, and this possibility should be further investigated.

It has been demonstrated in the beetle *Leptinotarsa decemlineata* that the dsRNase activity can be somewhat overcome by using RNAi to knockdown the enzymes that degrade dsRNA, but only to a point (Spit et al. 2017). A better avenue for overcoming this particular obstacle may be modifying the delivery system for the dsRNA or modifying the dsRNA itself. One promising option is the packaging of dsRNAs in biologically inert polymers, such as chitosan. The creation of dsRNA-chitosan nanocomplexes is relatively simple, and has been successful in knockdown of genes such as *vestigial* in *Aedes aegypti* (Mysore et al. 2013; Zhang et al. 2015; Ramesh Kumar et al. 2016). Another approach involves turning a pool of extremely efficient siRNAs into a larger, “self-dicing” dsRNA construct via disulfide bonds and complexing them within polyethylenimine nanospheres (Lee et al. 2010). In both of these methods, the polymer package protects the double-stranded RNA within from degradation by extracellular nucleases (Ragelle et al. 2013). More complex packaging approaches are being investigated in human clinical research to specifically target oncogenes in tumors, and some of these methods could be adapted to insects to target cells expressing particular receptors (Davis et al. 2010).

If, however, the dsRNA is being degraded not in the hemolymph of *E. carinata* but by non-Dicer enzymes within the cells, a more sophisticated approach is necessary. The endoribonucleases of concern are those that recognize double-stranded RNA preferentially. Chemical modification of the passenger strand of dsRNAs (but not the guide strand, which is the part loaded into the RNA-induced silencing complex) may be sufficient to evade intracellular degradation (Kanasty et al. 2012; Alagia and Eritja 2016). The results of these
future investigations may even yield data useful to researchers outside of the insect morphology and development community, and both the packaging approach and the modification approach are inexpensive enough that it is advisable to “just try it” (to paraphrase Scott et al. (2013)).

In this study, we have begun the work necessary to begin answering some of the unresolved questions around treehopper development, metamorphosis, and acquisition of adult form. We have established a year-round colony of *E. carinata* and have made substantial progress towards developing a working protocol for RNA interference in the nymphal stage. We have produced promising—though limited—results for functional assays of candidate developmental genes.

**Acknowledgements**

Many of the *E. carinata* specimens that were used to found the year-round greenhouse colony were collected by Hannah Ralicki. Michelle Deering and Adam Chiu assisted in colony maintenance. Funding for this project was provided by grants from the University of Connecticut Department of Ecology and Evolutionary Biology and the Connecticut Museum of Natural History.

**Author contributions**

Anticipated authors of a manuscript based on this chapter are C. R. Fisher and Elizabeth L. Jockusch. CRF and ELJ conceived the study and designed the experiments. CRF developed the injection protocol, performed the experiments, scored specimens, analyzed the data, and wrote the manuscript. ELJ edited the manuscript.
Literature cited


Smith, F. W., and E. L. Jockusch. 2014. Hox genes require homothorax and extradenticle for


Swevers, L., K. Ioannidis, M. Kolovou, A. Zografidis, V. Labropoulou, D. Santos, N.


### Table 1

Initial primers designed and used to amplify fragments of candidate genes from template cDNA. Degenerate primers developed using CODEHOP have a degenerate “core” (in lower case) with a non-degenerate “clamp” (in upper case).

<table>
<thead>
<tr>
<th>name</th>
<th>gene</th>
<th>Primer (5’ – 3’)</th>
<th>f/r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecar'TyrH-1</td>
<td>TH</td>
<td>TGAAGACCCCAAGATAGAAGAGA</td>
<td>F</td>
</tr>
<tr>
<td>Ecar'TyrH-2</td>
<td>TH</td>
<td>TCCTGAGAGAACTGAGCGAAG</td>
<td>R</td>
</tr>
<tr>
<td>Ecar'ebonyF1</td>
<td>e</td>
<td>CACTACCGTCACAGCAATTCAT</td>
<td>F</td>
</tr>
<tr>
<td>Ecar'ebonyR1</td>
<td>e</td>
<td>GTATAGCCGCCTCATATCTGGAG</td>
<td>R</td>
</tr>
<tr>
<td>Ecar'apF1</td>
<td>ap</td>
<td>GCTTCTCCCCGAGACGGTAAyathytytgya</td>
<td>F (outer)</td>
</tr>
<tr>
<td>Ecar'apF2</td>
<td>ap</td>
<td>TCTCCGCCACCCAGCTsgtbatgmvge</td>
<td>F (inner)</td>
</tr>
<tr>
<td>Ecar'apR1</td>
<td>ap</td>
<td>TTCTGGAAACCACACCTGCarnacnekytt</td>
<td>R (outer)</td>
</tr>
<tr>
<td>Ecar'apR2</td>
<td>ap</td>
<td>CCGTCTTTCTGCGACAGCTgyttnarctcyt</td>
<td>R (inner)</td>
</tr>
<tr>
<td>Ecar'scrF1</td>
<td>Scr</td>
<td>TGTCCTCCTACCAGTCTTGyGaaywsnytngc</td>
<td>F (outer)</td>
</tr>
<tr>
<td>Ecar'scrF2</td>
<td>Scr</td>
<td>ATGATGGACTACCTGGGAtncaynsnwc</td>
<td>F (inner)</td>
</tr>
<tr>
<td>Ecar'scrR1</td>
<td>Scr</td>
<td>GACATGTGGGTAGTGGAATActtancttact</td>
<td>R (outer)</td>
</tr>
<tr>
<td>Ecar'scrR2</td>
<td>Scr</td>
<td>CATCTTGTGCTCTTCTTctccayttctnnc</td>
<td>R (inner)</td>
</tr>
<tr>
<td>Ecar'scrF3</td>
<td>Scr</td>
<td>GCAGATCTACCCGTTGGATGaarmngntnca</td>
<td>F (inner)</td>
</tr>
<tr>
<td>Ecar'scrF4</td>
<td>Scr</td>
<td>GAACGCCAACGCGaracnaarmg</td>
<td>F (inner)</td>
</tr>
<tr>
<td>Ec'hth_F25</td>
<td>hth</td>
<td>GATTTCCTGTCGCCACTTTG</td>
<td>F</td>
</tr>
<tr>
<td>Ec'hth_R795</td>
<td>hth</td>
<td>TCTTCTCAAACAAATCCGACA</td>
<td>R</td>
</tr>
</tbody>
</table>
Table 2 T7 promoter-tagged primers used to generate template cDNA for dsRNA transcription.

<table>
<thead>
<tr>
<th>name</th>
<th>gene</th>
<th>primer (5' - 3')</th>
<th>f/r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ecar'T7apF1</td>
<td>ap</td>
<td>taatacgactcactataggGCCACTTTCCCTACCCCTCTACG</td>
<td>F</td>
</tr>
<tr>
<td>Ecar'T7apR1</td>
<td>ap</td>
<td>taatacgactcactataggGTGCATTCAAGTCTAGGTTAGCAG</td>
<td>R</td>
</tr>
<tr>
<td>Ecar'dll_t7_F5</td>
<td>Dll</td>
<td>taatacgactcactataggGACTCCACACTTTTCGTTTGG</td>
<td>F</td>
</tr>
<tr>
<td>Ecar'dll_t7_R5</td>
<td>Dll</td>
<td>taatacgactcactataggGGCTCCAACAACATCCCTACAAC</td>
<td>R</td>
</tr>
<tr>
<td>Ecar'scrT7_F1</td>
<td>Scr</td>
<td>taatacgactcactataggTGTAGTGTTAGGGGATAACG</td>
<td>F</td>
</tr>
<tr>
<td>Ecar'scrT7_R1</td>
<td>Scr</td>
<td>taatacgactcactataggCTGACAGAACGCCAGATCAA</td>
<td>R</td>
</tr>
<tr>
<td>T7Ecar'TH5</td>
<td>TH</td>
<td>taatacgactcactataggAGCCTGACCTGGACATGAAC</td>
<td>F</td>
</tr>
<tr>
<td>T7Ecar'TH6</td>
<td>TH</td>
<td>taatacgactcactataggGGCCAAGTAACTCGTGGATG</td>
<td>R</td>
</tr>
<tr>
<td>Ecar'T7_eF1</td>
<td>e</td>
<td>taatacgactcactataggGGTGCTCTGTGGTAAAGGTTCC</td>
<td>F</td>
</tr>
<tr>
<td>Ecar'T7_eR1</td>
<td>e</td>
<td>taatacgactcactataggTGTAAGACCGAAAAACATCATC</td>
<td>R</td>
</tr>
<tr>
<td>T7EC'hth_F</td>
<td>hth</td>
<td>taatacgactcactataggGCCAGCGCATTCAGTCTCCTAC</td>
<td>F</td>
</tr>
<tr>
<td>T7EC'hth_R</td>
<td>hth</td>
<td>taatacgactcactataggGCGAGGAGACGGAGGAGTAA</td>
<td>R</td>
</tr>
</tbody>
</table>

Table 3 Summary of injection cohort sizes, adult survival, and lethality of *E. carinata* RNAi versus controls. The P-values for mortality are from a logistic regression of each paired experiment (negative control versus experimental group). Cohort = number of 4th instar nymphs injected. N = number of surviving adults. Mortality = 1 – (N / cohort).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Conc. (μg/μL)</th>
<th>Exp. Cohort</th>
<th>Mortality</th>
<th>N</th>
<th>Neg. Control Cohort</th>
<th>Mortality</th>
<th>N</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ec'Th</td>
<td>5.5</td>
<td>30</td>
<td>40%</td>
<td>18</td>
<td>28</td>
<td>36%</td>
<td>18</td>
<td>0.74</td>
</tr>
<tr>
<td>Ec'eb</td>
<td>5.5</td>
<td>14</td>
<td>29%</td>
<td>10</td>
<td>10</td>
<td>30%</td>
<td>7</td>
<td>0.94</td>
</tr>
<tr>
<td>Ec'Scr</td>
<td>5</td>
<td>28</td>
<td>82%</td>
<td>5</td>
<td>14</td>
<td>50%</td>
<td>7</td>
<td>0.0325*</td>
</tr>
<tr>
<td>Ec'ap</td>
<td>5.5</td>
<td>41</td>
<td>44%</td>
<td>23</td>
<td>44</td>
<td>41%</td>
<td>26</td>
<td>0.78</td>
</tr>
<tr>
<td>Ec'Dll</td>
<td>6</td>
<td>55</td>
<td>65%</td>
<td>19</td>
<td>26</td>
<td>42%</td>
<td>15</td>
<td>0.0493*</td>
</tr>
<tr>
<td>Ec'hth</td>
<td>5</td>
<td>57</td>
<td>47%</td>
<td>30</td>
<td>27</td>
<td>33%</td>
<td>18</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Table 4 Summary of pigment pathway RNAi results: gene *Ec’ebony*, P-value result from logistic regression.

<table>
<thead>
<tr>
<th>Group</th>
<th>Cohort</th>
<th>N</th>
<th>Mortality</th>
<th>Phenotypic (black)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ec’e</em></td>
<td>15</td>
<td>10</td>
<td>33%</td>
<td>3</td>
<td>0.0569</td>
</tr>
<tr>
<td>dsGFP</td>
<td>10</td>
<td>7</td>
<td>30%</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5 Summary of pigment pathway RNAi results: gene *Ec’tyrosine hydroxylase*. P-value result from logistic regression.

<table>
<thead>
<tr>
<th>Group</th>
<th>Cohort</th>
<th>N</th>
<th>Mortality</th>
<th>Phenotypic (pale)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ec’TH</em></td>
<td>30</td>
<td>21</td>
<td>30%</td>
<td>14</td>
<td>0.003*</td>
</tr>
<tr>
<td>dsGFP</td>
<td>29</td>
<td>21</td>
<td>28%</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1 Results of RNAi targeting elements of the pigment pathway in *E. carinata*. **A, B.** Negative control male individual (injected with injection buffer following the protocol) showing typical *E. carinata* coloring. Overall color is dark brown. **B** shows detail of the legs, which are typically a pale tan color with dark tips on the spines of the third pair of legs (jumping legs). **C, D.** Female specimen showing phenotype resulting from knockdown of *Th*. Some black pigment is apparent, but the overall color is much lighter than in the wild type. **D** shows detail of the legs, which have much lighter T3 spines than in the wild type. **E, F.** Male specimen showing phenotype resulting from RNAi against *eb*. Very little yellow/tan pigment is apparent, so individual is much darker overall, with light speckles in across the helmet. **F** shows detail of the legs, which have black patches and almost entirely black femurs, atypical of wild-type *E. carinata*. 
Figure 2 Conserved domains of predicted Dicer-2 proteins. A. predicted DCR-2 from *E. carinata* reference transcriptome. The predicted protein contains the functional domains known to be required for cleavage of exogenous dsRNA in the RNAi silencing pathway, in the correct order. B. Comparison to DCR-2 from *Drosophila*. C. Comparison to DCR-2 from *Tribolium*. Conserved domains were detected by the ScanProSite web tool from ExPasy. GenBank accession numbers: *D. melanogaster* NP_001286540.1; *T. castaneum* NP_001107840.1
Figure 3 Conserved domains of predicted Argonaute-2 proteins. A. Predicted AGO-2 from *E. carinata* reference transcriptome. The protein contains the PAZ and PIWI functional domains. B., C-D. Functional domains of *Drosophila* and *Tribolium* AGO-2 proteins. *Tribolium* has two copies of ago-2.

GenBank accession numbers: *D. melanogaster* NP_648775.1; *T. castaneum* AGO-2a NP_001107842.1, *T. castaneum* AGO-2b NP_001107828.1
Figure 4 Gene tree constructed from amino acid sequences for proteins in the DCR family. 
*E. carinata* is predicted to possess both DCR-1 and DCR-2 based on the *E. carinata* reference transcriptome. The fruit fly protein Drosha, in the *Dicer* family, was included as an outgroup. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Dayhoff matrix based method and are in the units of the number of amino acid substitutions per site.

GenBank accession numbers: *D. melanogaster* Drosha NP_477436.1, Dcr-2 NP_0011286540.1, *Dcr-1* NP_524453.1, *T. castaneum* Dcr-2 NP_001107840.1, Dcr-1 XP_008199045.1
Figure 5 Gene tree constructed from amino acid sequences for predicted proteins in the AGO family. Based on the *E. carinata* reference transcriptome, treehoppers are expected to have an orthologous gene to *Drosophila* and *Tribolium ago-2*, the protein product of which is known to be required for RNA-induced silencing through the exogenous dsRNA pathway. *E. carinata* is also predicted to possess orthologs of *ago-1*, *ago-3*, and *aub*. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Dayhoff matrix based method and are in the units of the number of amino acid substitutions per site.

GenBank accession numbers: *D. melanogaster* Ago-2 Q9VUQ5, Ago-1 Q32KD4, Ago-3 Q7PLK0, Aub O76922.1; *T. castaneum* Ago-2a NP_001107842.1, Ago-2b NP_001107828.1, Ago-1 KYB26000.1, Ago-3 EFA02921.1, Aub XP_015837420.1
**Figure 6** Gene tree constructed from predicted dsRNA binding proteins LOQ and R2D2. E. carinata is predicted to possess a gene orthologous to r2d2, whose protein product is required for loading siRNA into the RISC in the exogenous dsRNA gene silencing pathway. The other clade of proteins, LOQ proteins, are involved in the miRNA silencing pathway. The evolutionary history was inferred using the Neighbor-Joining method. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Dayhoff matrix based method and are in the units of the number of amino acid substitutions per site.

GenBank accession numbers: *D. melanogaster* *loqs* Q4TZM6; r2d2 Q2Q0K7; *T. castaneum* *loqs* XP_966668.1, r2d2 NP_001128425.1, c3po XP_015835139.; *Nilaparvata lugens* *loqs* XP_022189835.1, r2d2 XP_022185715.
Figure 7 Schematic of a portion of the melanin synthesis pathway in insects, based on Liu et al. (2014) and Ferguson et al. (2011). The gene Tyrosine hydroxylase codes for a phenol oxidase (TH) required to convert the amino acid tyrosine into a melanin precursor. The gene ebony codes for an enzyme that converts dopamine and beta-alanine into a yellow pigment precursor. Knockdown of Th should reduce melanin production and give a light phenotype. Knockdown of eb should prevent conversion of dopamine into yellow pigment, and produce a dark phenotype.
Figure 8 Gel electrophoresis results of PCR products amplified from cDNA made from three pigment-pathway RNAi experimental specimens (318B, 319C, 319D) and three negative controls (317B, 317C, 317D). Left, amplification of a fragment of *Ec*’*Sex-combs reduced* (*Scr*) on all six templates indicates that RNA from extraction was intact (positive control). Right, amplification of *Ec*’*Th* is successful for negative controls but unsuccessful for experimentals, indicating successful depletion of the target mRNA.
RNA interference targeting *Ec ’Dll* resulted in a patchy reduction of distal portions of appendages with low penetrance. **A.** negative control male showing normal T1 and T2 legs. *E. carinata* have three tarsomeres in a tarsus that ends pretarsal claws. **B.** Male specimen from *Ec ’Dll* RNAi showing deleted tarsus and diminutive pretarsal claws. Arrowhead points to affected leg.
E. carinata forewings have a heavily sclerotized costal-subcostal region that bears strong structural similarity to the helmet. In order to rule out the possibility that helmet tissue and wing tissue transcriptional similarity was due to the forewing taking on aspects of the helmet’s patterning, rather than vice versa, we re-ran the differential expression analysis for E. carinata leaving forewings out.

The topology of the hierarchical clustering dendrogram remains the same. Branch support for all nodes that separate tissue types is 100. The similarity between treehopper helmets and wings is not due to the exoskeleton-like characteristics of the forewings.
Cluster dendrogram with AU/BP values (%)

Distance: euclidean
Cluster method: complete
Appendix B: TPM Normalization Between Species

library(knitr)
hook_output = knit_hooks$get('output')
knit_hooks$set(output = function(x, options) {
  # this hook is used only when the linewidth option is not NULL
  if (!is.null(n <- options$linewidth)) {
    x = knit:::split_lines(x)
    # any lines wider than n should be wrapped
    if (any(ncol(x) > n)) x = strwrap(x, width = n)
    x = paste(x, collapse = '\n')
  }
  hook_output(x, options)
})

Scaling TPM between two species with different numbers of annotated transcripts

based on Musser & Wagner (2015), JEZ:B

TPM, transcripts per million mapped, is a way of normalizing RNAseq data that accounts for differences in library size by scaling the abundance of a transcript (the “counts”) to the total number of transcripts assumed to be present in the transcriptome. In short, TPM = count for transcript “i” / total number of annotated transcripts * a scaling factor.

Necessarily, TPM from one species does not map to the TPM of another species if their transcriptomes are of different sizes, which they almost certainly are, so the values for the species with the smaller transcriptome will be inflated relative to the species with the larger transcriptome. According to Musser & Wagner, these values can be rescaled by calculating a scaling factor $\alpha$:

$$\alpha = 10^{-6} \sum_{j=1}^{N_1} tpm(A_j)$$

Where $N_1$ = the number of transcripts for species B (the smaller set), $j$ = all the transcripts in the set, and $tpm(A_j)$ is the transcripts per million for species A for each of the genes $j$ in the set.

Here is how I interpret this to work in my transcriptomes for *Entylia carinata* and *Homalodisca vitripennis*.

1) Read in the un-normalized TPM values from RSEM/edgeR.

```r
library("dplyr")
options(stringsAsFactors = FALSE)
Ecar.TPM <- read.table("D:/Cera Fisher/Google Drive/Treehoppers/ResearchFiles/RNASeq/...colnames(Ecar.TPM)
```

## [1] "X"    "ECA_Abd"  "ECA_Ovi"  "ECB_Abd"
## [5] "ECC_Abd"  "ECC_Ovi"  "ECEF_Abd"  "ECEF_Eye"
## [9] "ECEF_Leg"  "ECEF_Meso"  "ECEF_Ovi"  "ECEF_Pro"
## [13] "ECEF_Wing2" "ECEF_Wing3" "ECFisC_Abd" "ECFisC_Eye"
## [17] "ECFisC_Leg" "ECFisC_Meso" "ECFisC_Ovi" "ECFisC_Pro"
colnames(Ecar.TPM)[1] <- "ECid"

Hvit.TPM <- read.table("D:/Cera Fisher/Google Drive/Treehoppers/ResearchFiles/RNASeq/...colnames(Hvit.TPM)

colnames(Hvit.TPM)[1] <- "HVid"

Ofas.TPM <- read.table("D:/Cera Fisher/Google Drive/Treehoppers/ResearchFiles/...colnames(Ofas.TPM)

2) Calculate α

HvitN1 <- as.numeric(length(Hvit.TPM$HVid))
# 19,126

EcarN2 <- as.numeric(length(Ecar.TPM$ECid))
# 19,975

OfasN3 <- as.numeric(length(Ofas.TPM$OFid))
# 19,811

sumN2.EcarTPM <- (sum(Ecar.TPM[,2:42]))/41

sumN3.OfasTPM <- (sum(Ofas.TPM[,3])) # Ofas T1S1 is inflated. :

sumN3.OfasTPM <- (sum(Ofas.TPM[,2:12])/11) # Ofas T1S1 is inflated. :

# Sum of TPM for any given sample should, by definition, be 1,000,000
# The average TPM for Ecar is the sum divided by the number of transcripts.
Ec.avg.TPM <- sumN2.EcarTPM/EcarN2

# Getting the sum of Ecar TPM for the # of transcripts in Hvit's transcriptome
# i.e, multiplying the average times 19,126
sum.ECavgTPM.HvitN1 <- Ec.avg.TPM * HvitN1
sum.ECavgTPM.OfasN3 <- Ec.avg.TPM * OfasN3

# To get alpha, divide that amount by 1,000,000
alpha <- sum.ECavgTPM.HvitN1 * (10**(-6))

alpha

## [1] 0.9574969

beta <- sum.ECavgTPM.OfasN3 * (10**(-6))

This value for α, 0.957...etc is very close to the ratio of the smaller number of transcripts to the larger:

cHECKSUM <- HvitN1/EcarN2

cHECKSUM

## [1] 0.9574969

cHECKSUM - alpha

## [1] 7.006075e-09

Which perhaps should be expected, since those are the only values in this calculation that don’t cancel out.

Scaling Hvit.TPM, then, goes like this

HV_scaled <- Hvit.TPM[,1]
row.names(HV_scaled) <- Hvit.TPM[,1]
Hvit.TPM.scaled <- HV_scaled * alpha
dim(HV_scaled)

## [1] 19126 21

head(Hvit.TPM[,3])

## [1] 12.68 3.08 30.92 2.55 0.29 6.34

head(Hvit.TPM.scaled[,2])

## [1] 12.1410602 2.9490903 29.6058030 2.4416170 0.2776741 6.0705301

OF_scaled <- Ofas.TPM[,1]
row.names(OF_scaled) <- Ofas.TPM[,1]
Ofas.TPM.scaled <- OF_scaled * beta
dim(OF_scaled)

## [1] 15627 11

head(Ofas.TPM[,3])

## [1] 64.857 0.000 1.572 6.955 0.539 104.392

head(Ofas.TPM.scaled[,2])

## [1] 50.7394409 0.0000000 1.2298195 5.4410906 0.4216747 81.6687746

Multiplying by α results in our Hvit TPM numbers being just a little smaller, though it will matter a lot for some of the outrageously large numbers–

which(Hvit.TPM[,3] == max(Hvit.TPM[,3]))

## [1] 356
Hvit.TPM[,3]

Hvit.TPM.scaled[,2]

which(Ofas.TPM[,3] == max(Ofas.TPM[,3]))

Ofas.TPM[,3]

Ofas.TPM.scaled[,2]

Now, let’s save our scaled and size-factor normalized TPM to new files to use later.

head(Ofas.TPM.scaled)

Ofas_T3S2_TRINITY_DN14866_c4_g5_i1 0.2151402 50.7394409
Ofas_T3S2_TRINITY_DN19291_c0_g1_i1 0.0000000 0.0000000
Ofas_T3S3_TRINITY_DN10610_c0_g2_i1 0.0000000 1.2298195
Ofas_T1SCR_RNAi_1_TRINITY_DN22761_c0_g1_i1 1.5466623 5.4410906
Ofas_T3S2_TRINITY_DN15939_c0_g1_i1 1.8048305 0.4216747
Ofas_T1S1_TRINITY_DN5502_c0_g1_i1 52.8884956 81.6687746
Ofas_T1S2_TRINITY_DN14866_c4_g5_i1 47.5350258 6.9275136
Ofas_T3S2_TRINITY_DN19291_c0_g1_i1 0.5851813 0.1901057
Ofas_T3S3_TRINITY_DN10610_c0_g2_i1 1.6397593 1.0678776
Ofas_T1SCR_RNAi_1_TRINITY_DN22761_c0_g1_i1 7.6949773 6.9768002
Ofas_T3S2_TRINITY_DN15939_c0_g1_i1 1.8822809 0.7369529
Ofas_T1S1_TRINITY_DN5502_c0_g1_i1 96.7098109 119.8189948
Ofas_T1S3_TRINITY_DN14866_c4_g5_i1 2.943900 816.8622048
Ofas_T3S2_TRINITY_DN19291_c0_g1_i1 2.048917 0.9348818
Ofas_T3S3_TRINITY_DN10610_c0_g2_i1 0.0000000 0.0000000
Ofas_T1SCR_RNAi_1_TRINITY_DN22761_c0_g1_i1 1.667923 0.7080068
Ofas_T3S2_TRINITY_DN15939_c0_g1_i1 2.548042 0.8081447
Ofas_T1S1_TRINITY_DN5502_c0_g1_i1 59.788628 53.3500691
Ofas_T1S2_TRINITY_DN14866_c4_g5_i1 1.107776 0.4115045
Ofas_T3S2_TRINITY_DN19291_c0_g1_i1 1.456695 1.6953046
Ofas_T3S3_TRINITY_DN10610_c0_g2_i1 0.0000000 0.0000000
Ofas_T1SCR_RNAi_1_TRINITY_DN22761_c0_g1_i1 2.442428 0.0000000
Ofas_T3S2_TRINITY_DN15939_c0_g1_i1 1.884628 3.1879862
Ofas_T1S1_TRINITY_DN5502_c0_g1_i1 149.752425 146.8859975
Ofas_T2S2_TRINITY_DN14866_c4_g5_i1 149.752425 146.8859975
write.table(Ofas.TPM.scaled, "Ofas.90.isoforms.TPM.scaled.matrix", sep="\t")
write.table(Hvit.TPM.scaled, "Hvit.90.isoforms.TPM.scaled.matrix", sep="\t")
library(DESeq2)

## Loading required package: S4Vectors
## Loading required package: stats4
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
## The following objects are masked from 'package:parallel':
##
## clusterApply, clusterApplyLB, clusterCall, clusterEvalQ, clusterExport, clusterMap, parApply, parCapply, parLapply, parLapplyLB, parRapply, parSapply, parSapplyLB
## The following objects are masked from 'package:dplyr':
##
## combine, intersect, setdiff, union
## The following objects are masked from 'package:stats':
##
## IQR, mad, sd, var, xtabs
## The following objects are masked from 'package:base':
##
## anyDuplicated, append, as.data.frame, basename, cbind, colMeans, colnames, colSums, dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep, grepl, intersect, is.unsorted, lapply, lengths, Map, mapply, match, mget, order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank, rbind, Reduce, rowMeans, rownames, rowSums, sapply, setdiff, sort, table, tapply, union, unique, unsplit, which, which.max, which.min

## Attaching package: 'S4Vectors'
## The following objects are masked from 'package:dplyr':
##
## first, rename
## The following object is masked from 'package:base':

### expand.grid
### Loading required package: IRanges
### Attaching package: 'IRanges'
### The following objects are masked from 'package:dplyr':
### collapse, desc, slice
### The following object is masked from 'package:grDevices':
### windows
### Loading required package: GenomicRanges
### Loading required package: GenomeInfoDb
### Loading required package: SummarizedExperiment
### Loading required package: Biobase
### Welcome to Bioconductor
### Vignettes contain introductory material; view with 'browseVignettes()'. To cite Bioconductor, see 'citation("Biobase")', and for packages 'citation("pkgname")'.
### Loading required package: DelayedArray
### Loading required package: matrixStats
### Attaching package: 'matrixStats'
### The following objects are masked from 'package:Biobase':
### anyMissing, rowMedians
### The following object is masked from 'package:dplyr':
### count
### Loading required package: BiocParallel
### Attaching package: 'DelayedArray'
### The following objects are masked from 'package:matrixStats':
### colMaxs, colMins, colRanges, rowMaxs, rowMins, rowRanges
### The following objects are masked from 'package:base':
### aperm, apply

```r
colData <- read.table("SampleInformation_colData.txt", header=TRUE)
hvCol <- colData[25:45,]
hvTPM <- as.matrix(Hvit.TPM.scaled)
storage.mode(hvTPM) = "integer"
```
hv.dds <- DESeqDataSetFromMatrix(hvTPM, colData = hvCol, design = ~ Tissue + Pool)

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
hv.dds <- estimateSizeFactors(hv.dds)
hv.dds <- estimateDispersions(hv.dds)

## gene-wise dispersion estimates
## mean-dispersion relationship
## final dispersion estimates
hvScaledNorm <- as.data.frame(assay(hv.dds), normalized = TRUE)
write.table(hvScaledNorm, "Hvit_Scaled_SizeNormed_Integer_TPM.txt", sep="\t")

ECid <- Ecar.TPM[,1]
Ecar.TPM <- Ecar.TPM[,] - 1
Ecar.TPM <- Ecar.TPM[,c(6:13,14:21,31:38)]
rownames(Ecar.TPM) <- ECid
ec.dds <- as.matrix(Ecar.TPM)
eCol <- colData[1:24,]
storage.mode(ec.dds) = "integer"

ec.dds <- DESeqDataSetFromMatrix(ec.dds, colData = ecCol, design = ~ Tissue + Pool)

## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
ec.dds <- estimateSizeFactors(ec.dds)
ec.dds <- estimateDispersions(ec.dds)

## gene-wise dispersion estimates
## mean-dispersion relationship
## -- note: fitType='parametric', but the dispersion trend was not well captured by the
## function: y = a/x + b, and a local regression fit was automatically substituted.
## specify fitType='local' or 'mean' to avoid this message next time.
## final dispersion estimates
ecNorm <- as.data.frame(assay(ec.dds), normalized=TRUE)
write.table(ecNorm, "Ecar_SizeNormed_Integer_TPM.txt", sep="\t")
Appendix C: Linear discriminant analysis with MLSeq

```r
library(knitr)
hook_output = knit_hooks$get('output')
knit_hooks$set(output = function(x, options) {
  # this hook is used only when the linewidth option is not NULL
  if (!is.null(n <- options$linewidth)) {
    x = knitr:::split_lines(x)
    # any lines wider than n should be wrapped
    if (any(ncol(x) > n)) x = strwrap(x, width = n)
    x = paste(x, collapse = '\n')
  }
  hook_output(x, options)
})
```

Read in selected markers, filter the counts matrices.

```r
selectedMarkers <- read.table("1_Tuned_MergedCounts_SelectedMarkers_10000iter.txt")
unselected <- filter(Merged.Counts, !Merged.Counts$OrthoID %in% selectedMarkers$OrthoID))
```

Plot the *E. carinata* TPM vs. *H. vitripennis* TPM for genes identified as species markers.

991 transcripts that discriminate for species

This plot shows a trident pattern. Extreme outliers in expression along the axes indicate genes highly
expressed in one species and not expressed in the other.

**Plot** *E. carinata* TPM vs. *H. vitripennis* TPM of genes not selected as species markers.

6657 transcripts classified as unbiased

This plot does not have a trident pattern. Most genes with expression biased towards one species have been removed from the set.

```r
write.table(selectedMarkers, "MLSeq_PLDA_SelectedMarkers_991.txt")
write.table(unselected, "UnselectedTranscripts_6657.txt")
```

Let’s take a look at how the PCAs from my DESeq2 script (based on a tutorial by Jill Wegrzyn) look when they’re based on just this set of orthologs. For the sake of brevity, I ran the differential expression analysis and variance stabilizing transform separately, and will just load it here.

```r
vst <- read.table("HV_EC_Orthos-vst-transformed-counts.txt")
library("RColorBrewer")
library("gplots")
distsRL <- dist(t(vst))
mat <- as.matrix(distsRL)
rownames(mat) <- colnames(mat) <- with(colData(dds),
  paste(Tissue, Sample, sep=" : "))
#Or if you want conditions use:
#rownames(mat) <- colnames(mat) <- with(colData(dds),condition)
hmcol <- colorRampPalette(brewer.pal(9, "GnBu"))(100)
heatmap.2(mat, trace = "none", col = rev(hmcol),
  margin = c(13,13))
```
Hierarchical clustering still gives a deep split with *H. vitripennis* samples on one branch and *E. carinata* samples on the other. However, *E. carinata* abdominal samples are now clustering with *H. vitripennis* samples.

```r
# Principal components plot shows additional but rough clustering of samples
top <- rowVars(vst)
samples <- colData$Sample
tissue <- colData$Tissue
species <- colData$Species
pool <- colData$Pool

scores <- data.frame(pc$x, species)

pcaplot <- ggplot(scores, aes(x = PC1, y = PC2, col = (factor(species)), shape = factor(tissue))) +
  scale_shape_manual(values=c(6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18, 6,7,8,3,10,16,17,18))
```

205
Principal Components

% of variance explained by each principle component.

summary(pc)

## Importance of components:
## Standard deviation
## Proportion of Variance

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye</td>
<td>0.2213</td>
<td>0.1548</td>
<td>0.1356</td>
<td>0.09508</td>
<td>0.05346</td>
<td>0.04926</td>
</tr>
<tr>
<td>Component</td>
<td>Proportion of Variance</td>
<td>Cumulative Proportion</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------</td>
<td>-----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC1</td>
<td>0.2213</td>
<td>0.2213</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC2</td>
<td>0.3761</td>
<td>0.6067</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC3</td>
<td>0.5117</td>
<td>0.66019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC4</td>
<td>0.6067</td>
<td>0.70946</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC5</td>
<td>0.66019</td>
<td>0.74311</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC6</td>
<td>0.70946</td>
<td>0.76908</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC7</td>
<td>0.80926</td>
<td>0.82750</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC8</td>
<td>0.84508</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC9</td>
<td>0.92138</td>
<td>0.92922</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC10</td>
<td>0.94459</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC11</td>
<td>0.94910</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC12</td>
<td>0.97534</td>
<td>0.97820</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC13</td>
<td>0.97835</td>
<td>0.98093</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC14</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC15</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC16</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC17</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC18</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC19</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC20</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC21</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC22</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC23</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC24</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC25</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC26</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC27</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC28</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC29</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC30</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC31</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC32</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC33</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC34</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC35</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC36</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC37</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC38</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC39</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC40</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC41</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC42</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC43</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC44</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC45</td>
<td>0.97835</td>
<td>0.98350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

pcaplot <- ggplot(scores, aes(x = PC2, y = PC3, 
  col = (factor(species)),
  shape=(factor(tissue)))) + 
  scale_shape_manual(values=c(6,7,8,3,10,16,17,18,
                             6,7,8,3,10,16,17,18,
                             6,7,8,3,10,16,17,18,
                             6,7,8,3,10,16,17,18,
                             6,7,8,3,10,16,17,18,
                             6,7,8,3,10,16,17)) +
  geom_point(size = 5) +
  ggtitle("Principal Components") +
  scale_colour_brewer(name = "", palette = "Dark2") +
  theme(
    plot.title = element_text(face = 'bold'),
    legend.key = element_rect(fill = 'NA'),
    legend.text = element_text(size = 10, face = "bold"),
    axis.text.y = element_text(colour = "Black"),
    axis.text.x = element_text(colour = "Black"),
    axis.title.x = element_text(face = "bold"),
    axis.title.y = element_text(face = "bold"),
    panel.grid.major.x = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor.x = element_blank(),
    panel.grid.minor.y = element_blank(),
    )

pcaplot
```r
panel.background = element_rect(color = 'black', fill = NA)
)
picaplot

pcaplot <- ggplot(scores, aes(x = PC3, y = PC1,
    col = (factor(species)),
    shape = (factor(tissue)))) +
    scale_shape_manual(values = c(6, 7, 8, 3, 10, 16, 17, 18,
                               6, 7, 8, 3, 10, 16, 17, 18,
                               6, 7, 8, 3, 10, 16, 17, 18,
                               6, 7, 8, 3, 10, 16, 17, 18)) +
    geom_point(size = 5) +
    ggtitle("Principal Components") +
    scale_colour_brewer(name = "", palette = "Dark2") +
    theme(
        plot.title = element_text(face = 'bold'),
        legend.key = element_rect(fill = 'NA'),
        legend.text = element_text(size = 10, face = "bold"),
        axis.text.y = element_text(colour = "Black"),
        axis.text.x = element_text(colour = "Black"),
        axis.title.x = element_text(face = "bold"),
        axis.title.y = element_text(face = 'bold'),
        panel.grid.major.x = element_blank(),
        panel.grid.major.y = element_blank(),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_blank(),
        panel.background = element_rect(color = 'black', fill = NA)
    )
```
Across all three of the first three axes, the species signal still separates samples most deeply. However, the order of tissue types across the axes is conserved, except for the place of the pronotum tissues. In particular, note that the pronota for *E. carinata* is clustered with wings for *E. carinata* and *H. vitripennis*, and the pronota for *H. vitripennis* is clustered with mesonota and appendages.

Abdominal tissues and eye tissues dominate the difference between tissue types, but on different axes of variation. This complexity cannot be captured in hierarchical clustering and only becomes apparent in multidimensional scaling-type analyses.

**Code used to create PLDA classifier and separate species markers from other genes.**

```r
#~~~~ Cera Fisher (2018)
### MLSeq - finding species marker genes with machine learning

library("dplyr")
library("DESeq2")
library("MLSeq")

#Read in TPM matrix
MergedCounts <- read.delim("HV_EC_NewSamples.SingleCopyOrthogroups.MergedTPM.txt", sep="\t", header=TRUE)
```
colnames(MergedCounts)[c(2, 3)] <- c("HVid", "ECid")

c <- as.matrix(MergedCounts[,4:48])
storage.mode(c) = "integer"

class <- DataFrame(condition = factor(rep(c("Ecar", "Hvit"), c(24, 21))))
## Setting up a Class object for DESeq2

set.seed(2128)
vars <- sort(apply(c, 1, var, na.rm = TRUE), decreasing = TRUE)
data <- cts[ names(vars)[1:300], ]
GoodInd <- read.table("good_ind.txt", sep="\t")
ind <- as.vector(GoodInd)
ind <- ind[,1]
data.train <- as.matrix(data[, -ind] + 1)
data.test <- as.matrix(data[, ind] + 1)
class <- DataFrame(condition = class[-ind, ])
class <- DataFrame(condition = class[ ind, ])
c <- DESeqDataSetFromMatrix(countData = cts.train,
                           colData = class, design = formula(~condition))
featureData <- data.frame(gene=MergedCounts$OrthoID)
mcols(c) <- DataFrame(mcols(c), featureData)
mcols(c) <- DataFrame(mcols(c), data.frame(HVid=MergedCounts$HVid))
mcols(c) <- DataFrame(mcols(c), data.frame(ECid=MergedCounts$ECid))
ctrl.PLDA <- discreteControl(method="repeatedcv", number=30,
                           repeats=10000,
                           rho=23.89522,
                           parallel=TRUE)

fit.all.PLDA <- classify(c, method="PLDA", preProcess="deseq-vst",
control=discreteControl(ctrl.PLDA))

Markers <- selectedGenes(fit.all.PLDA)
Markers.Counts <- MergedCounts[Markers,]
write.table(Markers.Counts,
           "Tuned_MergedCounts_SelectedMarkers_10000iter.txt")