

**Genetic Determinants of the Development and Evolution of Drosophila
Pigmentation**

By

Abigail M. Lamb

A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular, and Developmental Biology)
In the University of Michigan
2021

Doctoral Committee:

Professor Patricia Wittkopp, Chair
Associate Professor Scott Barolo
Associate Professor Laura Buttitta
Professor Kenneth Cadigan

Abigail M. Lamb

abbylamb@umich.edu

ORCID iD: 0000-0001-5184-4180

Dedication

To everyone who started graduate school but, for any reason, didn't continue to the point of dissertation defense. I was lucky to have the support I had, otherwise I would not have made it to this point. I have seen too many wonderful, talented people pushed out of academic science, and they deserved so much better than what they experienced.

Acknowledgments

I would first like to thank Trisha Wittkopp, not only for her scientific and professional mentorship, but especially for her support, kindness, and understanding during (possibly too many) years of struggle and success – both at the bench and in life. Going to college in the first place felt like a stretch goal given the circumstances of my early life, and graduate school even now somehow sounds impossibly unlikely, even as I complete my dissertation. I truly believe that my ability to succeed in graduate school was dependent on finding an environment and culture where I could thrive, and the Wittkopp Lab became Home to an extent that is difficult to describe.

For fear of forgetting someone, I will not attempt to individually name all of the many, many labmates I spent time with both in and out of the lab. However, I will give special mention to Elizabeth Walker for being the reassuring and welcoming presence I needed when I showed up not even knowing how to PCR, Andrea Hodgins-Davis for being one of the kindest human beings I've ever met, Jen Lachowiec for having the best laugh and talking me into playing hockey, Mark Hill for all the chats over beverages and horror cinema, Jade Diaz for being extra and venting with me, Lisa Kim for helping me keep things weird, and Mo Siddiq even though he spends too much time giving me genuine compliments as a form of torture. Special thanks to Anati Azhar and Patty Simmer for allowing me the privilege of being your mentor through the emotional roller coaster of academic science research, and to Alicia Wang, Lisa Kim, Alisha John, and Elizabeth Walker for making everything work while simultaneously being delightful.

Outside of my home lab: Thank you to Scott Barolo, Laura Buttitta, and Ken Cadigan for serving on my thesis committee and always having excellent feedback and advice. In MCDB, I am grateful to all the helpful and supportive staff, the Graduate Student Council, and everyone who poured themselves into building the culture they want for themselves and those who come after them. Special shout out to Shyama Nandakumar for being a fellow chaotic softie with whom I could commiserate over drinks and houseplants after we said “yes” to too many things. Thank you to my collaborators, Jennifer Kennell, Zinan Wang, and Henry Chung, and my funding sources: The National Science Foundation Graduate Research Fellowship Program and The Michigan Predoctoral Training Program in Genetics.

I’ve had the privilege to form a family of widely varying degrees of genetic relatedness, especially Jennifer, Jon, Glenna, Dorris, Bob, Shannon, Laura, Maureen, Jessica, Meg, Anthony, Garrett, Abigail, Bella, and Kali. Thank you for getting me through the (many) rough patches and making it possible for me to keep going. Thank you to the science educators and mentors who helped me find my path: Cindy Bagwill, Jerry Dunaway, and Tony Frankino. Thank you to all the students I’ve had the privilege to teach: y’all were a joy. And lastly, there is no chance I could have done any of this without Michael, who occupies his own category and is the single most important person in my life.

And thank you to Strobel, because I wouldn’t have gotten in the door without him.

Table of Contents

Dedication	ii
Acknowledgments	iii
List of Tables	viii
List of Figures	x
List of Appendices	xiii
Abstract	xiv
Chapter 1: Introduction	1
Chapter 2: Tools and Strategies for Scarless Allele Replacement in <i>Drosophila</i> Using CRISPR/Cas9	30
Abstract	30
Introduction	31
Results	35
Discussion	41
Materials and methods	45
Acknowledgments	49

Abbreviations	49
References	50
Chapter 3: <i>ebony</i> Affects Pigmentation Divergence and Cuticular Hydrocarbons in <i>Drosophila americana</i> and <i>D. novamexicana</i>	56
Abstract	56
Introduction	57
Materials and Methods	60
Results and Discussion	67
Conclusions	80
Acknowledgements	83
References	83
Chapter 4: microRNAs Are Necessary Components of the Genetic Architecture Underlying Adult Cuticle Pigmentation in <i>Drosophila melanogaster</i>	88
Abstract	88
Introduction	89
Results and Discussion	91
Conclusion	113
Materials and Methods	114
References	117
Chapter 5: Conclusions and Future Directions	123
References	133

List of Tables

Table 1-1: Cases of phenotypic evolution attributed to differences in miRNA regulation in GePheBase.....	7
Table 4-1 Summary of phenotype data from all miRNA overexpression crosses.	95
Table 4-2. Summary of phenotypic effects of overexpression and competitive inhibition of all microRNAs included in competitive inhibition screen.	101
Supplementary Table S2-1: Oligonucleotides and primers used in reagent construction and molecular screening.	144
Supplementary Table S2-2: Genome editing rates in the first stage of the allele swap.	148
Supplementary Table S2-3: Genome editing rates in the second stage of the allele swap.	149
Supplementary Table S3-1: Oligonucleotides used to generate sgRNAs for CRISPR/Cas9 genome editing and <i>nos-Cas9-nos</i> transgene.....	156
Supplementary Table S4-1: Phenotypes and notes on all individual flies collected in miRNA overexpression screen.	166
Supplementary Table S4-2: Phenotypes and notes on all individual flies collected in miRNA competitive inhibition screen.....	167

Supplementary Table S4-3: Summary data from overexpression and competitive inhibition screens.....	168
Supplementary Table S4-4: List of pigmentation genes.....	169

List of Figures

Figure 1-1: miRNA deletions affecting phenotypes in <i>Caenorabditis elegans</i> and <i>Drosophila melanogaster</i>.	12
Figure 1-2: Schematic of biosynthetic pathway and genetic network underlying <i>Drosophila melanogaster</i> pigmentation, reproduced from Massey and Wittkopp 2016.	14
Figure 1-3: Looking under the lamppost.	21
Figure 2-1: Challenges for single-stage allele replacement strategies using CRISPR.	32
Figure 2-2: Schematic of marker-assisted, 2-stage allele swap within the <i>D. americana tan</i> transgene in <i>D. melanogaster</i>.	35
Figure 2-3: Representative pigmentation and fluorescence phenotypes of flies at each stage of the <i>tan</i> allele swap process.	39
Figure 2-4: Workflow for 2-stage marker assisted allele swap.	42
Figure 3-1: <i>Drosophila novamexicana</i> shows divergent body color within the virilis group.	59
Figure 3-2: <i>ebony</i> affects body, wing, and pupal pigmentation in <i>D. novamexicana</i> and <i>D. americana</i>.	69
Figure 3-3: CRISPR/Cas9-induced mutations created null alleles of the <i>D. novamexicana</i> and <i>D. americana ebony</i> genes.	72

Figure 3-4: Reciprocal hemizyosity testing shows effects of <i>ebony</i> divergence between <i>D. americana</i> and <i>D. novamexicana</i> on body pigmentation.	74
Figure 3-5: Cuticular hydrocarbons (CHCs) are affected by <i>ebony</i> and differ between <i>D. americana</i> and <i>D. novamexicana</i>.	77
Figure 3-6: <i>ebony</i> does not contribute to divergence of CHCs between <i>D. americana</i> and <i>D. novamexicana</i>.	79
Figure 4-1: Schematic of miRNA overexpression and competitive inhibition screens and scoring parameters.....	92
Figure 4-2: Pigmentation phenotypes resulting from overexpressing 153 miRNAs.	94
Figure 4-3: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen.	98
Figure 4-4: Images demonstrating miRNAs that are both sufficient to alter abdominal pigmentation and necessary for normal pigmentation development.	103
Figure 4- 5: Computationally-predicted seed matches between known pigmentation genes and miRNAs that showed reciprocal phenotypes when overexpressed versus competitively inhibited.	106
Figure 4-6: miR-8 coordinately regulates <i>ebony</i>, <i>black</i>, <i>bric a brac 1</i> and <i>bric a brac 2</i> to promote the production of dark pigments in female A6.....	112
Supplemental Figure S2-1: Chromatograms showing the sequence of the edited sites before and after each stage of the editing process.....	150

Supplemental Figure S3-1: Mutations in the <i>D. novamexicana</i> and <i>D. americana</i> <i>white</i> genes cause white-eyed phenotype.	152
Supplemental Figure S3-2: <i>nanos</i> promoter failed to drive expression of Cas9 in transgenic <i>D. novamexicana</i>.	153
Supplemental Figure S3-3: Mutation in the <i>D. novamexicana</i> <i>yellow</i> gene causes visible changes in body pigmentation.	154
Supplemental Figure S3-4: Uncropped and un-annotated western blot images and membranes.	155
Supplemental Figure S4-1: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen, male data, competitive inhibition crosses at 28°C.	160
Supplemental Figure S4-2: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen, female data, competitive inhibition crosses at 18°C.	162
Supplemental Figure S4-3: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen, male data, competitive inhibition crosses at 18°C.	164

List of Appendices

Appendix A: Supplemental Data for Chapter 2	138
Appendix B: Supplemental Data for Chapter 3.	152
Appendix C: Supplemental Data for Chapter 4.	159

Abstract

Understanding the relationships between DNA sequence and organismal phenotypes has been a central goal across a wide variety of biological subdisciplines throughout the history of genetics research. Studies of the developmental genetic mechanisms underlying phenotypes aim to increase our understanding of how genotypes encode phenotypes, while studies of the genetic basis of evolved differences inform our understanding of how changes in DNA sequence alter these developmental processes. While the fields of evolution and development have been mutually informative, our ability to interpret genetic sequences remains severely limited. In this dissertation, I investigate the genetic basis of phenotypic evolution and development using pigmentation in the genus *Drosophila* as a model system. In order to overcome technical limitations in identifying the genetic basis of evolved differences in *Drosophila* pigmentation, I adapted CRISPR/Cas9 genome editing techniques for precise DNA sequence replacement, developing methods designed for use in a wide variety of *Drosophila* species. Using CRISPR/Cas9, I demonstrate the role of the *ebony* gene in an evolved pigmentation difference between *Drosophila novamexicana* and *Drosophila novamexicana*, as well as in the development of cuticular hydrocarbons in these species. Finally, I describe the previously under-studied role of microRNAs in the development of *Drosophila melanogaster* pigmentation in a large-scale genetic screen, introducing a new class of regulators into the existing body of research into the genetic architecture underlying this development of this model trait.

Chapter 1: Introduction

The information required for the development and function of each organism is encoded within the DNA sequence of its genome. However, this DNA sequence on its own is inert, and must be expressed by the machinery of the cell in order for the functions of life to proceed. Thus, the genomic DNA of an organism must contain information that encodes the molecules that “read” the DNA, as well as the instructions for robust and tightly controlled deployment of all molecules used by the cell throughout its lifetime. Any given gene’s spatial, temporal, and environmentally contextual expression patterns are controlled by a multitude of complex molecular systems. Activation or repression of transcription depends upon the coordination of chromatin structure, epigenetic markers, complex transcription factor binding, DNA-DNA contacts through three-dimensional looping, and many more phenomena to control the initial RNA output of a gene. While regulation of transcription is complex on its own, there are many more layers of regulation beyond initial RNA production, including the processing, splicing, nuclear export, and selective degradation of RNA molecules, followed by further regulation at the level of translation and post-translational processing. Cell-to-cell differences in regulatory networks determine the gene expression profiles that specify cell identity, response to environmental cues, and all other biological processes.

In addition to their necessity for all biological functions, the genetically encoded developmental programs that underlie all phenotypes also provide the raw material for natural variation and evolution. In much the same way as differences in the expression patterns of genes lead to the differentiation of individual cell types within an organism, many of the phenotypic differences we see between individuals of the same species and between taxa across the tree of life result in large part from changes in gene regulation over evolutionary time.

Perspectives and insight from decades of research connecting genotypes and phenotypes

Understanding the relationships between DNA sequence, gene expression, genetic network structure, and phenotypic diversity is a long-standing goal of biological research across a wide variety of disciplines. The first studies linking genotypes to phenotypes used “forward genetics” methods to identify regions of an organism’s genome where sequence differences correlate with a phenotypic difference (Sturtevant 1913). This method of inquiry begins by identifying a phenotype of interest that shows some heritable variation, either as naturally segregating variants or as lab-generated mutant phenotypes induced through random DNA mutagenesis. Identifying regions of the genome where different alleles correlate with different phenotypes, either through classical genetic mapping or through modern methods like genome-wide association studies (GWAS), uncovers candidate sequences that affect the phenotype’s development (Sturtevant 1913; Ikegawa 2012). However, identifying regions of DNA sequence that affect a phenotype does not provide mechanistic insight - we may know *that* a difference in this sequence is important for a particular phenotype, but we don’t yet know *how* the information encoded in the region affects the functions or activities of specific genes. Once forward genetic studies reveal genomic regions and sequence variants that affect a phenotype, “reverse genetic” methods may be used to alter DNA sequence or manipulate genes in these candidate regions in order to explore how these sequences and genes affect the phenotype (Hardy et al. 2010).

A great body of research is dedicated to identifying and characterizing the genetic differences that underlie phenotypes that show great diversity across populations and species. Over 1700 alleles that have been implicated in phenotypic differences across the all eukaryotes over the last ten years are compiled in Gephebase, a database that catalogues the supporting evidence, mutation characteristics (if known), taxonomic data, and references supporting these genotype-phenotype interactions (Courtier-Orgogozo et al. 2019). This large volume of work

informs our hypotheses as we explore the ways in which existing genetic and developmental mechanisms underlying phenotypes may affect evolutionary trajectories.

One recurring theme throughout these studies suggests that physiological traits and genes at terminal points of regulatory networks are more likely to evolve by changes in coding sequence, while morphological traits, tissue-specific traits, and changes in genes situated higher up in genetic networks appear to preferentially evolve by differences in non-coding sequences (David L. Stern and Orgogozo 2008; Carroll 2005). In addition, biases in our ability to identify the genetic basis of some changes more readily than others are apparent. In the case of coding versus non-coding sequence changes, for example, protein-coding sequences have historically been easier to identify in most cases, as coding sequences generally have straightforward boundaries and much of the time we can tell which variants are likely to affect the function of the sequence because of our understanding of the genetic code (David L. Stern and Orgogozo 2008). However, as we develop new and more sensitive methods of genetic analysis, changes in non-coding sequences that alter the regulation of one or more genes seem to contribute to a larger portion of case studies. Increasing amounts of data began support the cis-regulatory hypothesis, which posits that changes to non-coding regulatory sequences may be evolutionarily favored over coding changes, particularly in cases where the focal gene acts in multiple tissues, developmental stages, or biological processes (Carroll 2005; David L. Stern and Orgogozo 2008). Changes to regulatory sequences, such as enhancers, can alter the output of a gene in a restricted set of developmental or environmental circumstances, while coding changes are usually universal across tissues, developmental stages, and environmental conditions. The evolving ratio of coding to non-coding examples underscores the importance of ascertainment bias and technical limitations in the types of mechanisms we are likely to discover.

Overcoming limitations: barriers to finding genetic causes of phenotypic evolution

While our understanding of the mechanisms of phenotypic evolution is much more complete than it once was, there are still many areas where our understanding is limited. For one, a long-standing issue is that many studies stutter to a halt after identifying candidate genetic regions or sequences where allelic differences associate with a change in phenotype, but before resolving mechanistic details or specific causative mutations (Abiola et al. 2003). This is often attributable to a variety of factors, including difficulties performing reverse genetics in more than a handful of historically well-developed model organisms, large amounts of sequence variation over evolutionary time making it hard to know which differences are important, structural issues such as rearrangements (inversions) interfering with mapping, and reproductive isolation across the evolutionary distances at which many changes are observed, to name a few. Intriguingly, there is also the possibility that our discoveries are constrained not only because of technical limitations, but also because we may not be primed to look in the right places for underlying mechanisms.

Ascertainment bias can occur due to both biological and technical reasons. As described above, there are cases of biological bias where some changes in DNA sequences are easier to interpret, such as protein coding sequences relative to non-coding regulatory sequences. Similarly, single alleles with large phenotypic effects are easier to identify than large numbers of small-effect mutations scattered across many loci with a large cumulative effect (Rockman 2012). However, biases in interpretation are likely pervasive, as we are restricted by our historical knowledge of the genetics within our model phenotypes and organisms. Even in studies that use a relatively unbiased approach, such as genetic mapping and GWAS, we are still biased in our interpretation of the results. Indeed, upon identifying a candidate region where allelic differences correlate with a change in phenotype, our next logical step is generally to look for nearby genes with characterized functions that may relate to our focal phenotype. In other words, if a candidate region contains or is in close proximity to a gene that has been shown to have a promising function, one is likely to perform follow-up experiments focused on querying the role of that gene. Otherwise, one might

conclude that they have reached a “dead end” of negative results if no promising genes are nearby or if follow-up experiments on promising candidate genes fail to produce convincing evidence that said genes are involved in the phenotypic change under examination. This is sometimes referred to as the “streetlight effect” or “looking under the lamppost”, referring to a joke wherein a man continuously searches for his missing wallet under a single streetlight even though he doesn’t know where he dropped it because “the light’s better here” (Freedman 2010). If the likely candidate genes in a region have been exhaustively studied without producing convincing evidence of a causative role, it is difficult to know how to proceed in areas of a candidate region that are not yet “illuminated” by previous research suggesting a promising function. Compounding this issue is the tendency of “negative results” to often go unpublished, leaving the scientific community unaware of the potential of these poorly understood loci as projects either accumulate evidence for what *isn’t* happening or are eventually abandoned. We are restricted to making interpretations and prioritizing lines of investigation based on our knowledge of which genes have an established relationship with our focal genotype.

Overcoming limitations: getting out from under “the lamppost”

As an example of types of mutations we might possibly miss due to limited perspective or biased expectations, changes in post-transcriptional regulation by microRNAs (miRNAs) are theoretically good candidates for cis-regulatory evolution, but are thus far rarely found to underlie phenotypic differences. This class of small, non-coding RNAs canonically act as negative post-transcriptional regulators by binding to the 3’ untranslated regions (3’ UTRs) of messenger RNAs in a sequence-specific manner, directing the RNA-induced silencing complex to prevent translation of the targeted gene into protein (Bartel 2004, 2018). As described above, cis-regulatory sequences such as promoters and enhancers are thought to prevent deleterious pleiotropic effects by limiting a mutation’s effects to a specific developmental stage or

tissue. Similarly, changes to the expression patterns of miRNAs or to miRNA binding sites in the 3' UTRs of their target genes could be expected to have the same benefits, since miRNAs are only expected to exert their effects on the mRNA transcripts of their target genes where the expression patterns of both miRNA and target overlap (K. Chen and Rajewsky 2007). However, of the 1700+ entries in GePheBase, at the time of this writing only six cases are attributed to changes in miRNA expression or miRNA-target interactions (Table 1-1) (Courtier-Orgogozo et al. 2019).

The dearth of case studies identifying miRNA-based mechanisms underlying phenotypic differences could easily be assumed to result from biological reality, drawing the conclusion that changes in miRNA regulatory networks are not common routes for phenotypic evolution despite their attractive potential for modularity. In contrast to this interpretation, it could be that our methods, presumptions, and limited information are holding us back from discovering cases where miRNA regulation is responsible for phenotypic differences. Upon closer examination, the few case studies supporting a role of miRNAs in the evolution of phenotypes provide some compelling reasons to suspect that post-transcriptional regulation by miRNAs may be a more common part of the evolutionary tool kit than is currently appreciated.

Gephe ID	Species	Trait	Taxonomic level of comparison	Affected gene(s)	Short description	Reference(s)
GP00000168	<i>Ovis aries</i> (sheep)	Hindquarter muscularity with "polar overdominance" mode of inheritance. Known as "callipyge" phenotype.	Domestic breeds/cultivars	<i>Dkl1*</i> (maternally silenced) <i>miR-379/miR-544</i> cluster* (paternally silenced) *Both within the imprinted <i>Dkl1-Dio3</i> region	<i>Dkl1</i> expression in muscle is sufficient to increase muscle mass. One or more miRNAs in the nearby maternally expressed <i>miR-379/miR-544</i> cluster repress <i>Dkl1</i> in wild-type expression context. In callipyge individuals, the CLPG mutation is inherited from the male parent, and paternal <i>Dkl1</i> is misexpressed in postnatal skeletal muscle in the absence of maternal miRNAs, leading to muscular hypertrophy.	(Georges et al. 2003; Cao et al. 2015; Takeda et al. 2006; Fraking et al. 2002; Murphy et al. 2006)
GP00000661	<i>Drosophila melanogaster</i> (fly)	Trichome patterns on adult legs	Intraspecific variation	<i>mir-92a</i>	Changes in cis-regulation cause expanded expression of <i>mir-92a</i> , which represses <i>shavenoid</i> , leading to a larger region of trichome-free "naked" cuticle.	(Anif et al. 2013)
GP00000695	<i>Ovis aries</i> (sheep)	Muscularity	Domestic breeds/cultivars	<i>GDF8</i> (myostatin)	A point mutation in the 3' UTR of the gene encoding myostatin (<i>GDF8</i>) creates a target site for two miRNAs, miR-1 and miR-206, which are both highly expressed in skeletal muscle. This causes translational repression of myostatin and muscular hypertrophy.	(Clap et al. 2006)
GP00000813	<i>Oryza sativa</i> (rice)	Grain yield	Domestic breeds/cultivars	<i>OsSPL14</i>	"Ideal plant architecture" underlying increased grain yield was mapped to a point mutation in <i>OsSPL14</i> that perturbs a OsmiR156 binding site. OsmiR156 regulation of <i>OsSPL14</i> was found to underlie the development of traits that affect grain yield.	(Jiao et al. 2010)
GP00001314, GP00001315	<i>Hordeum vulgare</i> (barley)	Pollen shedding	Domestic breeds/cultivars	<i>Cly1</i>	In a comparison of 274 barley cultivars, cleistogamy (closed flowering) associates perfectly with a point mutation that disrupts a miR172 target site in the <i>Cly1</i> gene.	(Nair et al. 2010)
GP00002247	<i>Gallus gallus</i>	Body weight	Domestic breeds/cultivars	<i>miR-15a-16</i>	A 54-bp deletion upstream of <i>miR-15a-16</i> significantly associates with increased body weight in GWAS. This deletion decreases <i>miR-15a-16</i> expression, possibly leading to increased cell proliferation.	(Jia et al. 2016)

Table 1-1: Cases of phenotypic evolution attributed to differences in miRNA regulation in GePheBase (On previous page. Data from Courtier-Orgogozo et al 2019).

Case studies suggest we may be missing important regulatory mechanisms in phenotypic evolution

Take, for example, the discovery of the role of a miRNA, *miR-92a*, in the evolution of intraspecific variation in trichome patterning on the legs of *D. melanogaster* flies (Arif et al. 2013). Previous studies had found that changes in cis at the gene *Ultrabithorax (Ubx)* were responsible for interspecific differences in leg trichome patterning between *D. melanogaster* and *Drosophila simulans*, making this gene a promising candidate for variation in this trait within *D. melanogaster* (*D. L. Stern 1998*). Genetic mapping efforts identified a single quantitative trait locus (QTL) that accounted for approximately 90% of the variation in this trait between the two *D. melanogaster* populations under investigation. While this QTL was located on the same chromosome as *Ubx*, it was physically distant from the *Ubx* locus, and follow-up experiments directly contrasting the phenotypic effects of the two populations' *Ubx* alleles showed no effect on leg trichome patterning (Arif et al. 2013). Upon searching the QTL for other potential candidate genes, no protein-coding genes were found that had a known or suspected role in trichome development. However, the QTL did contain *miR-92a*, which had recently been demonstrated to affect trichome development in a study wherein a library of 180 miRNAs were over-expressed and screened for effects on a set of phenotypes (Schertel et al. 2012; Arif et al. 2013). After following up on *miR-92a* as a candidate gene, Arif et. al. found *miR-92a* expression differed notably between the two mapping populations, determined that the difference in expression was likely caused by a cis-regulatory difference, and presented evidence that strongly suggested *miR-92a* affects this phenotype by repressing the function of *shavenoid*, a known target of the transcription factor *shavenbaby*. For context, *shavenbaby* is a notorious “hotspot” gene underlying multiple instances of trichome evolution in the *Drosophila* genus that is thought to hold an evolutionarily tractable position within the trichome developmental

network (David L. Stern and Frankel 2013; Arif, Kittelmann, and Mcgregor 2015; David L. Stern and Orgogozo 2008).

One can only speculate as to whether Arif et. al. would have identified *miR-92a* as the gene responsible for intraspecific variation in trichome patterning if a miRNA screen had not implied a promising role of *miR-92a* in the development of this phenotype only the year before, but this case study effectively illustrates how even “unbiased” approaches to genetic mapping can be easily influenced by follow-up experiments that carry intrinsic bias due to limitations in our understanding of a trait’s genetic architecture. After all, time and resources are limiting factors in our ability to follow-up on unbiased mapping and association studies, meaning it would be unrealistic and unwise for investigators to perform follow-up experiments on every identifiable gene or regulatory element in a candidate region, regardless of our understanding of the roles of these genes and sequences. Therefore, we must be somewhat dependent on existing data to prioritize follow-up experiments, which unfortunately may prevent us from finding the specific mutations and mechanisms contributing to a region’s effects on a phenotype of interest in some cases.

Another surprising discovery of trait evolution via post-transcriptional regulation by miRNAs occurred in the search for the genetic basis of a hypertrophic muscularity phenotype in domestic sheep selected bred for meat production (Clöp et al. 2006; Georges et al. 2006). Genetic mapping between hypermuscular Texel sheep and regularly muscled Romanov sheep identified a QTL that explained 20-33% of the difference in muscularity between the parental populations. Excitingly, this QTL contained the *GDF8* gene, also known as *myostatin*, which is a well-established genetic “hotspot” that has been implicated in evolved differences in musculature or racing performance in several species of domestic animals including cows, sheep, dogs, pigs, and horses (Aiello, Patel, and Lasagna 2018; Courtier-Orgogozo et al. 2019). In all mechanistically characterized cases, loss or reduction of Myostatin function leads to increased muscle mass or improved racing performance, often through changes to protein-coding sequence (Aiello, Patel, and Lasagna 2018; Courtier-Orgogozo et al.

2019). However, in the case of the hypermuscular Texel sheep, there were no differences in *GDF8* coding sequence and no significant difference in *GDF8* mRNA levels in muscles between the two mapped breeds. Upon further examination, Clop et al. found that a polymorphism in the 3' UTR of *GDF8* created a recently described miRNA recognition motif in the Texel allele, which could reasonably lead to reduced translation of *GDF8* mRNA in the presence of any of three known miRNAs (*miR-1*, *miR-206*, and *miR-122a*) and thus reduced amounts of Myostatin protein (Xie et al. 2005). Further, the authors note that *miR-1*, a miRNA predicted to target the binding motif present in the 3' UTR of the Texel *GDF8* allele, had recently been implicated in skeletal and cardiac muscle development in mice (Zhao, Samal, and Srivastava 2005). Follow-up experiments demonstrated reduced Myostatin protein levels in Texel sheep muscle relative to Romanov sheep muscle and provided evidence for direct allele-specific repression of the Texel allele, but not the Romanov allele, of the *GDF8* 3' UTR by *miR-1* and *miR-206*, both of which are highly expressed in sheep skeletal muscle (Clop et al. 2006).

Interestingly, while the hypermuscularity of Texel sheep was being investigated, other groups were continuing ongoing investigations into the genetic basis of the “callipyge” (derived from a Greek word meaning “beautiful buttocks”) phenotype in domestic sheep selected for exceptionally meaty hindquarters (Cockett et al. 1999; Takeda et al. 2006; Cockett et al. 1994). The callipyge phenotype shows an unusual inheritance pattern of polar overdominance, in which only heterozygotes inheriting the CLPG mutant single-nucleotide polymorphism from the male parent display the characteristically muscular rumps associated with the mutation (Cockett et al. 1999, 1996). After years of study, the most recent theory for a mechanism underlying this strange mode of non-Mendelian inheritance involves differential regulation of a genetically imprinted region containing three protein-coding genes and many non-coding RNAs, where the protein-coding genes are silenced in the maternal allele, but the non-protein-coding genes are silenced in the paternal allele (Gao et al. 2015). All the genes in this region show ectopic expression in postnatal skeletal muscle when associated in *cis* with the CLPG allele, with one protein-coding gene in particular, *Dlk1*

(Delta-like 1 homolog), hypothesized to cause muscular hypertrophy (Charlier et al. 2001; Davis et al. 2004; Takeda et al. 2006). Intriguingly, evidence from genetic manipulations suggests *Dlk1* mRNA is targeted by one or more miRNAs in the neighboring miRNA cluster (Gao et al. 2015). Therefore, when the paternal allele carries the ectopically expressed CLPG variant, paternal *Dlk1* expression occurs in postnatal skeletal muscle while the paternal miRNA cluster is silenced, and the maternal allele of *Dlk1* is silenced by imprinting while the maternal miRNA allele is expressed in the wild-type pattern, which excludes postnatal skeletal muscle expression. It is hypothesized that miRNA/*Dlk1* RNA co-transcription in postnatal skeletal muscle may silence the phenotypic effects of ectopic *Dlk1* expression in homozygous CLPG/CLPG individuals via miRNA-mediated translational repression of *Dlk1*, leading to the polar overdominance inheritance pattern of the callipyge phenotype (Gao et al. 2015).

Expanding our focus: unexpected mechanisms may be more important than we realize

After the Texel sheep study demonstrated the potential role of polymorphic miRNA-target interactions in the evolution and development of muscle in a livestock species, many studies have since investigated the role of miRNAs in agriculturally important traits. Studies in sheep, ducks, cows, pigs, goats, chickens, and trout have all found evidence suggesting miRNA involvement in intraspecific variation in a wide variety of traits including body weight, muscle yield, muscle quality, wool characteristics, egg-laying, milk production, litter size, ear size, and sperm quality to name a few, all of which cite the Texel sheep study by Clop et al. 2006 (C. Chen et al. 2018; Paneru et al. 2017; Ali et al. 2019; Wu et al. 2017; Qiu et al. 2020; An et al. 2016; Jia et al. 2016).

In reviewing these case studies, it is clear that our ability to determine the genetic causes of differences in phenotypes depends on historical knowledge and mechanistic understanding of the model system under investigation. In the *Drosophila* leg trichome and Texel sheep muscularity cases described above, the authors' stated rationale in investigating the role of miRNAs in their phenotype of interest rested on previous

studies that had either implicated a particular miRNA in the developmental process underlying their phenotype (Arif et al. 2013; Clop et al. 2006), or identified candidate genes by characterizing the genetic architecture underlying their focal trait through genetic manipulations (Gao et al. 2015). It seems reasonable to speculate that many later studies investigating the roles of miRNAs in the development of agriculturally desirable traits might not have been pursued if the role of miRNAs in hypermuscular sheep had not been demonstrated previously. After all, in early studies where miRNAs were systematically mutated in *Caenorhabditis elegans*, most did not show any obvious phenotype, leading many to conclude that changes in miRNA-mediated regulation were unlikely to have discernable effects (Miska et al. 2007). This notion of relative unimportance of miRNAs persists, despite the fact that later studies in more organisms, assaying a wider variety of phenotypes, and under a variety of environmental conditions revealed that individual miRNAs often hold important developmental roles in the regulation of gene expression and the development of animal phenotypes (Bartel 2018). For instance, while the miRNA deletion study by Miska et al in 2007 only found 8 out of 87 *C. elegans* miRNA deletions affected any of the 8 phenotypes assayed, a 2014 study in *D. melanogaster* found that, of 95 miRNA deletions, 76 affected at least one of 11 phenotypes tested (Figure 1-1) (Miska et al. 2007; Y. W. Chen et al. 2014). It appears increasingly likely that the role of miRNAs in

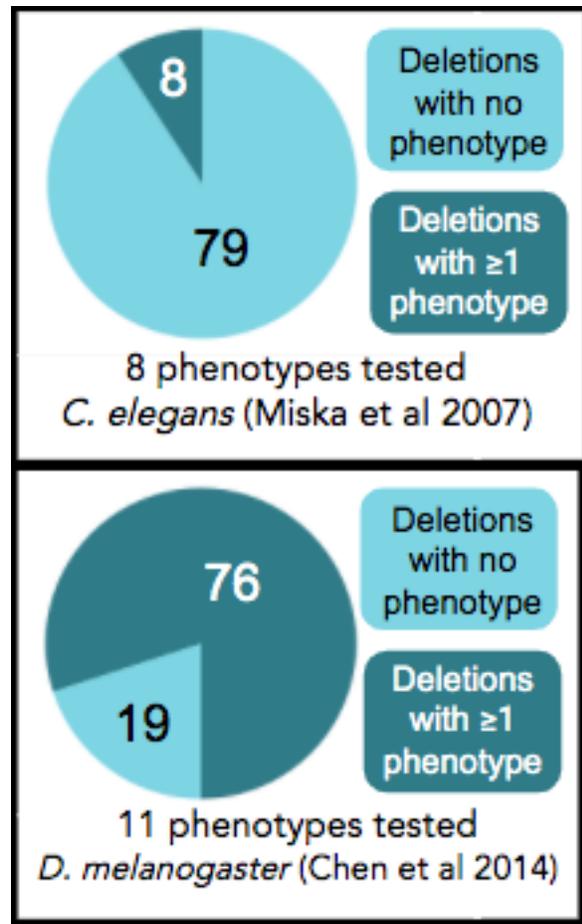


Figure 1-1: miRNA deletions affecting phenotypes in *Caenorhabditis elegans* and *Drosophila melanogaster*. Counts of miRNA deletions affecting 0 phenotypes (light cyan), and deletions affecting 1 or more phenotypes (dark cyan) in two studies screening miRNA deletion mutants for phenotypic effects in *Caenorhabditis elegans* (top) and *Drosophila melanogaster* (bottom).

regulatory networks, and by extension in phenotypic evolution, may be underappreciated (K. Chen and Rajewsky 2007). The studies described in this manuscript illustrate the importance of establishing a thorough understanding of the genetic architecture underlying model traits in order to facilitate new insights into the wide variety of biological mechanisms that may serve as evolutionary paths to diverse phenotypes. One of the great benefits of model organisms and model traits is our ability to build upon a well-established knowledge base describing the developmental genetics underlying these traits, so we would be well-served by actively seeking out unknown regulators or mechanisms in unbiased, exploratory studies in order to gain new insights into the ways in which these model traits evolve.

Insect pigmentation as a model system for the study of genotype-phenotype interactions

Coloration has been the one of the most productive model phenotypes in the history of research connecting genotypes to phenotypes across a wide variety of taxa, second only to xenobiotic resistance and accounting for 18% of case studies documented in Gephabase (Kronforst et al. 2012; Hoekstra 2006; Sobel and Streisfeld 2013; Wittkopp and Beldade 2009; Courtier-Orgogozo et al. 2019). Insect pigmentation has been a particularly well-studied model system. Pigmentation is extremely diverse across arthropods, and the genetic and biosynthetic mechanisms underlying insect pigmentation development and regulation have been, and continue to be, very thoroughly studied (Figure 1-2) (True 2003; Wittkopp and Beldade 2009; Andersen 2010). In addition, pigmentation is an ecologically relevant and biologically important trait, with roles in mate recognition, camouflage, thermoregulation, and water balance (True 2003; Wittkopp and Beldade 2009). Consistent with pigmentation's roles in multiple biological processes, the genes involved in pigmentation synthesis are pleiotropic, playing multiple roles across developmental stages and tissues (Wittkopp and Beldade 2009; Takahashi 2013). For instance, *ebony* and *tan* genes encode enzymes that are crucial in the melanin biosynthesis pathway, but act on different substrates in the nervous system where they are involved in neurotransmitter recycling (Borycz et al. 2002; True et al. 2005; Aust et al. 2010). Furthermore, *ebony* and *tan*

have recently been shown to affect cuticular hydrocarbons, hydrophobic lipids on the adult cuticle, which are involved in water balance, chemical communication, and mate recognition in insects (Chung and Carroll 2015; Chung et al. 2014; Jonathan H. Massey et al. 2019). Genes involved in pigmentation are also regulated in a sex-specific manner in many species to produce sexually dimorphic coloration in adult animals (Kopp et al. 2000; Monteiro 2015). Tightly regulated spatial, temporal, and sex-specific expression of genes involved in pigmentation development is vital to allow the formation of many complex traits, making it a fascinating system for the study of development, and evolution, and the intersection of the two (Rebeiz and Williams 2017; Wittkopp and Beldade 2009).

Much of our understanding of genotype-phenotype relationships in insect pigmentation is built upon a foundation of developmental and evolutionary studies in *D.*

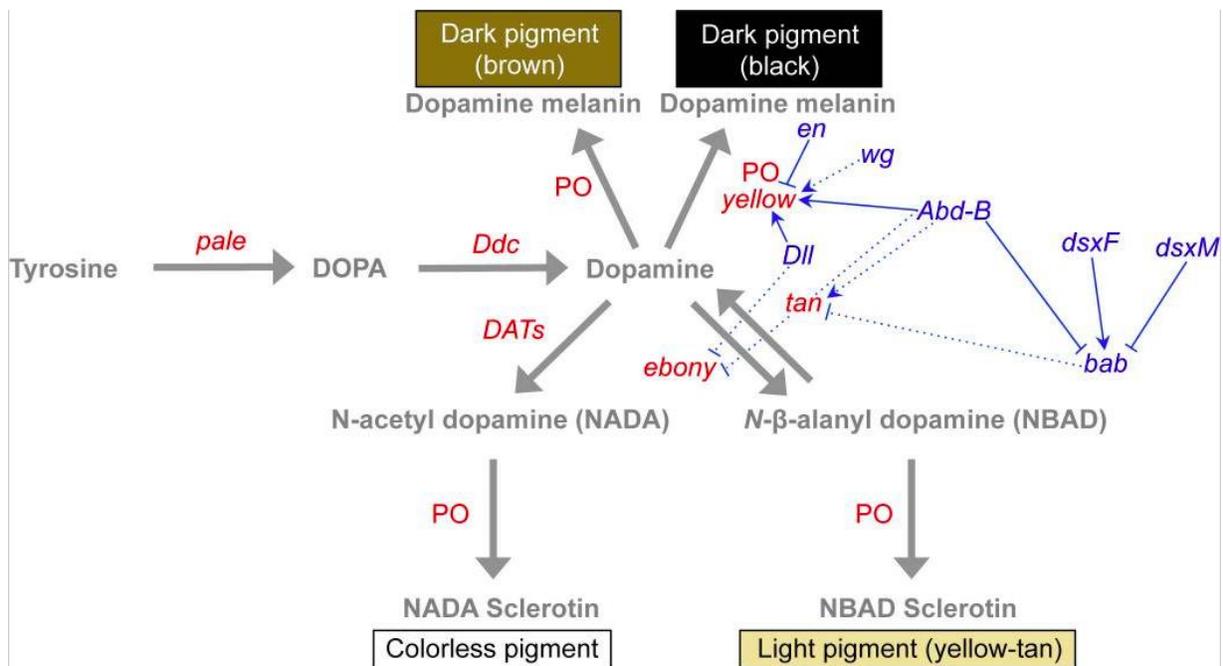


Figure 1-2: Schematic of biosynthetic pathway and genetic network underlying *Drosophila melanogaster* pigmentation, reproduced from Massey and Wittkopp 2016. Metabolites are represented in gray, genes directly involved in pigment synthesis are in red, and genes involved in the regulation of pigmentation are in blue. Solid and dotted blue arrows represent direct and indirect regulatory connections, respectively. Arrowheads represent positive regulation, while nail heads represent negative regulation.

melanogaster and other species in the *Drosophila* genus (Wittkopp, Carroll, and Kopp

2003; J. H. Massey and Wittkopp 2016; Rebeiz and Williams 2017). Dozens of case studies have identified, to varying degrees of specificity, the genetic causes of differences in body and wing pigmentation both within and between *Drosophila* species; at the time of this writing, every case where the molecular basis of a difference has been identified, it has resulted from changes in cis-regulation of genes known to affect pigmentation synthesis or development (J. H. Massey and Wittkopp 2016). However, there are still many cases where the identified genetic mechanisms only account for a portion of the phenotypic difference observed between populations, or where phenotypic differences have been mapped to QTL but have not yet been connected to the action of a particular gene (J. H. Massey and Wittkopp 2016). Difficulties in identifying the specific molecular basis of these differences are likely a combination of technical limitations and incomplete understanding of the genetic architecture underlying pigmentation development.

Technological advances grant access to what we've been missing

In addition to the constraints imposed by our knowledge of the genetic architecture of a trait's development and regulation, investigations into the genetic basis of natural variation face a variety of technical limitations. Even within well-established model species like *D. melanogaster* and *C. elegans*, studies of natural variation intrinsically require the use of strains that differ in the focal phenotype, and these strains are unlikely to be the common laboratory stocks that serve as the genetic backgrounds for most deletion collections and transgenic resources. Studies of phenotypes that evolved over longer evolutionary periods, such as interspecific differences, are not only limited by the lack of transgenic resources and mutants, but also must contend with reproductive isolation, chromosomal rearrangements, and large amounts of sequence divergence.

Take, for example, the derived light yellow body color of *Drosophila novamexicana* relative its sister species, *Drosophila americana*, and the rest of the Virilis species group, all of which are brown to black in color (Novitski and Ashburner 1976). Initial efforts to map the genetic basis of this change in body color revealed an X chromosome effect and three autosomal QTL which contributed to this phenotype (Wittkopp et al. 2003). The X-linked candidate genes known at the time were not found to affect pigmentation, and only one of the 3 autosomal QTL contained a known candidate gene: *ebony*, which encodes a synthetase that converts dopamine molecules to N-beta alanyl dopamine (NBAD), a yellow pigment precursor (Figure 1-2) (Wittkopp et al. 2003). Unfortunately, numerous inversions between the mapping populations severely limited the resolution attainable through recombination mapping since recombination within an inversion causes lethal dicentric chromosomes, leaving the *ebony*-containing QTL linked to ~15% of the second chromosome (Wittkopp et al. 2003). Later fine-scale mapping, introgression, and transgenic analysis showed that *tan*, a hydrolase which converts NBAD to dopamine in the inverse of *ebony*'s catalytic function, contributed to this pigmentation difference, and further suggested that non-coding differences within the first intron of *tan* were responsible (Wittkopp et al. 2009). However, while this second set of experiments also found that a large QTL linked to *ebony* showed a very strong effect on pigmentation, the inversion still prevented any further resolution, leaving the possibility any one of the many genes within the QTL or any combination of these genes could be responsible for the region's effect on pigmentation (Wittkopp et al. 2009). Differences in *ebony* expression between these species were detected at both the protein and mRNA levels, but these expression differences continued to provide only correlative evidence, since the phenotypic effect of the difference in *ebony* expression could not be observed while *ebony* remained linked to a large inverted region (Wittkopp et al. 2003; Cooley et al. 2012).

While the effects of *tan* were able to be separated from other loci on the X chromosome, and were in fact mapped to a <3kb region of non-coding sequence with a finite number of sequence differences between the mapping populations, efforts to identify the specific mutations responsible for the difference have yet to be realized.

Transgenic work in *D. melanogaster tan* null mutants successfully demonstrated that genetic rescue by *D. novamexicana* and *D. americana* alleles of *tan* yielded lighter and darker abdominal body pigmentation, respectively (Wittkopp et al. 2009; John et al. 2016). However, these experiments also demonstrated the difficulties and limitations of comparing the effects of foreign alleles in an evolutionarily distant genetic background. The phenotypic effects of the *D. novamexicana* and *D. americana* alleles could distinguished in some cases when both transgenes were inserted at the same genomic location, but the difference between these alleles was indistinguishable when the transgenes were integrated at some loci, despite the fact that both alleles were expressed and could rescue pigmentation in the *tan* mutant host (John et al. 2016). Further experiments comparing chimeric transgenes with the intronic candidate region swapped between the species' alleles did not yield reliable results (A. John and A. Lamb, data not shown), suggesting that assessment of these alleles in the trans genetic context of a distantly related species may not be a feasible method for identifying the mutations that cause these alleles to produce different pigmentation phenotypes. While transgenic analysis is possible in non-model drosophilids, it relies on random integration, meaning that the effects of transgenes cannot be controlled for position effects. Controlling for position effects of randomly integrated transgenes is difficult in any experiment, and in the case of these specific alleles, we already have evidence that chromosomal location has a strong impact on our ability to interpret their effects on pigmentation even when they can be integrated into a known landing site. Between the limitations of transgenic methods and the coarseness of genetic mapping on account of chromosomal inversions, technological limitations stalled investigation of the genetic basis of light body color in *D. novamexicana*.

Ideally, direct genetic manipulations at the native *ebony* and *tan* loci within *D. americana* and *D. novamexicana* were the most desirable and easy-to-interpret experimental strategies. For example, in the case of pigmentation changes due to evolution at *tan*, the region implicated through mapping is limited to the first intron of *tan* and contains 57 single-nucleotide changes and 19 insertions or deletions (Wittkopp et al. 2009). The most desirable approach would be to replace portions of *D.*

novamexicana sequence in this region with homologous *D. americana* sequence and observe pigmentation to see where within this region the causative changes reside. Identification of the specific mutation(s) responsible for changes in *tan* expression would allow for follow-up experiments to query the functional differences between alleles to identify the molecular mechanism(s) by which this gene's expression evolved. Until recently, these types of manipulations have only been possible through transgenic methods, often in distantly related species, which suffer from the difficulties described here. Excitingly, with the advent of CRISPR genome editing, genetic manipulations became theoretically plausible in any species which could be transformed through embryonic injection (Jinek et al. 2012; Doudna and Charpentier 2014; Barrangou and Horvath 2017). Indeed, CRISPR has been used to modify the genomes of a wide variety of taxa, including arthropod species spanning nine taxonomic orders (Sun et al. 2017; Gratz, Cummings, et al. 2013).

Unfortunately, organisms outside the small set of extensively developed model species face many challenges that can hinder the efficiency and practicality of genetic manipulations, even with the advent of species-agnostic methods such as CRISPR. While genome editing in *D. melanogaster* has become remarkably efficient through the use of resources such as transgenic lines that express the molecular machinery necessary for CRISPR in specific tissues and developmental stages, similar resources have not yet become widely available in non-model organisms - although transgenic resources for CRISPR genome editing are gradually becoming available in a wider variety of species (Port, Muschalik, and Bullock 2015; David L. Stern et al. 2017). Efforts at genetic manipulations in most non-model species often only produce mutants at low frequencies, likely due to a combination of microinjection/rearing difficulties and the need to inject all components needed for editing rather than relying on stably integrated transgenic sources (Sun et al. 2017). Screening for low-frequency editing events in non-model species is laborious and time-consuming, especially in cases where the desired mutations are not expected to produce a clearly visible phenotype in a heterozygous state (Kane et al. 2017). Methods which rely on the insertion of selectable markers to overcome this challenge may leave undesirable sequence

insertions, deletions, or changes in addition to the desired changes, making the results less straightforward to interpret (Bier et al. 2018). To overcome this challenge, I devised an allele-replacement strategy that implements a visible, selectable marker and is designed to facilitate the insertion of a series of alleles without the need for tedious molecular screening and without leaving behind undesired mutations in addition to the desired edits (Lamb, Walker, and Wittkopp 2017). This method is described in Chapter 2 of this dissertation.

In addition to allele replacement techniques, CRISPR genome editing is frequently used to generate loss-of-function mutations of target genes (Gratz, Wildonger, et al. 2013). The phenotypic effects of two divergent alleles of a gene can be compared via a “reciprocal hemizygosity test” so long as (a) loss-of-functions mutants can be generated in the two populations of interest and (b) the two populations can be crossed and produce viable offspring (David L. Stern 2014). This method is well-suited for separating the effects of a single gene from the effects of linked loci when searching for the genetic basis of a phenotypic difference since the reciprocal hemizygosity test allows the comparison of phenotypes between hybrids that are genetically identical except for the parental origin of their functional copy of the gene of interest (David L. Stern 2014). By generating *ebony* mutant alleles in both *D. novamexicana* and *D. americana* using CRISPR genome editing, we were able to demonstrate the effects of divergence at *ebony* on pigmentation using reciprocal hemizygotes, providing the first direct phenotypic evidence for the role of *ebony* in the evolution of light yellow body color in *D. novamexicana* (Lamb et al. 2020). In Chapter 3 of this dissertation, I describe the generation of genome editing methods and resources for *D. novamexicana* and *D. americana*, the production of the first CRISPR-generated mutants in these species, and the effects of *ebony* on pigmentation divergence between *D. americana* and *D. novamexicana*. We also show that *ebony* affects the abundance of cuticular hydrocarbons in both these species, suggesting that the recently described role of *ebony* in cuticular hydrocarbon production may be evolutionarily conserved across a wide number of *Drosophila* species (Jonathan H. Massey et al. 2019; Lamb et al. 2020).

Turning on more lights: improving our search by finding new regulators

While many of the enzymes involved in the synthesis of cuticle pigments in *Drosophila* have been known for some time, efforts to flesh out the genetic regulation and development of *Drosophila* pigmentation in a more detailed manner began to take on momentum as it was first established as a model system for the evolution of development around the turn of the 21st century (Wright et al. 1982; Wright 1987; Biessmann 1985; Walter et al. 1991; Kopp et al. 2000; Wittkopp, True, and Carroll 2002; Wittkopp, Carroll, and Kopp 2003). Over the past two decades, many genes have been found to affect the development of pigmentation in *Drosophila* through genetic mapping, GWAS, and genetic screens (J. H. Massey and Wittkopp 2016; Dembeck et al. 2015; Kalay et al. 2016; Rogers et al. 2014). Genetic screens using RNAi to assess the effects of hundreds of transcription factors on pigmentation have identified many previously unknown regulators of pigmentation, though in most cases the specific roles and regulatory connections of these transcription factors remain unknown (Kalay et al. 2016; Rogers et al. 2014). GWAS was used to identify sequence variants that associate with pigmentation variation among over 200 inbred *D. melanogaster* strains of the *Drosophila* Genome Reference Panel (DGRP), which led to the discovery of 17 genes previously unknown to affect pigmentation (Dembeck et al. 2015). These types of unbiased studies not only add to our understanding of the complex genetic architecture underlying pigmentation development, but will undoubtedly clarify and facilitate studies aiming to find the genetic mechanisms underlying pigmentation differences. The transcription factor RNAi screens and within-species GWAS efforts described above illustrate the utility of searching broadly for new regulators, but both are, like all studies, necessarily limited in scope. The RNAi screens are focused specifically on transcription factors to the exclusion of all other types of genes, and the GWAS of DGRP lines was limited to finding genes that have variant alleles that affect pigmentation within the study population. More experiments with different limitations can be used to search widely for unknown regulators and are likely to illuminate our understanding of this study system even further.

Like most model traits, the role of post-transcriptional regulation in both the development and evolution of insect pigmentation is not well understood. To my knowledge, the only example of a miRNA affecting insect pigmentation is the miRNA *miR-8*, which has been shown to affect pigmentation in the dorsal abdomen of female *D. melanogaster* (Kennell et al. 2012). This effect was only discovered upon mutating *miR-8* while investigating its role in Wingless signaling and finding that the dark melanization of the posterior-most dorsal abdominal segments of female flies was noticeably reduced in *miR-8* mutants (Kennell et al. 2012). The developmental roles, network interactions, and expression patterns of most *D. melanogaster* miRNAs are not well characterized (Lucas and Raikhel 2013). *Drosophila* pigmentation could theoretically be well-suited as a model system in which to study the regulation, interactions, and functions of miRNAs in development, evolution, plasticity, and sexual dimorphism, to name a few possibilities, but this could only be the case if this trait's development involves regulation by miRNAs. In Chapter 4 of this dissertation, we demonstrate the role of several miRNAs in *D. melanogaster* pigmentation, first by over-expressing a library of miRNAs and observing their effects on pigmentation, and then following up on candidates discovered in this screen.

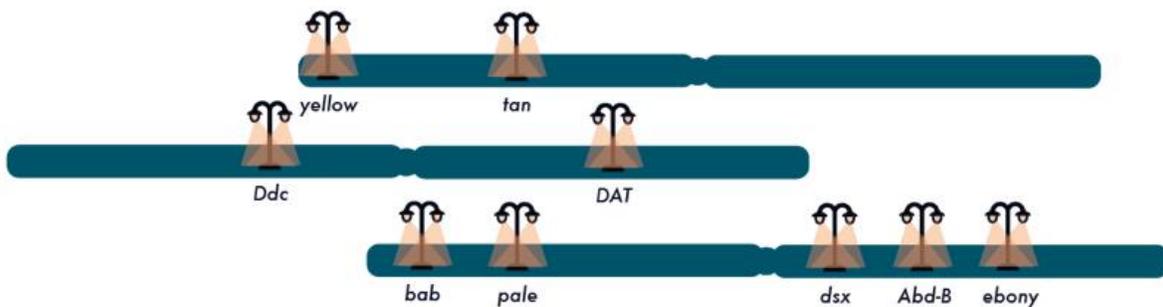


Figure 1-3: Looking under the lamppost. Bars represent *D. melanogaster* chromosomes (from top: X, 2, 3 – not to scale). Cartoon lampposts represent genes that have been found to underlie differences in pigmentation phenotypes within and/or between species in the *Drosophila* genus. Without understanding more about the genetic architecture controlling the development of pigmentation, our ability to identify the genes causing differences in pigmentation is mostly limited to a few well-characterized options.

The studies in chapters 2-4 of this dissertation open many questions and possibilities for follow up. The diversity of *Drosophila* pigmentation may provide an excellent opportunity to study the evolution of miRNA regulatory networks. The increasing tractability of genetic manipulations of multiple *Drosophila* species with the

advent of new technologies also clears the path to more complex and detailed studies of gene expression, evolution, and the evolution of gene expression. In chapter 5, I describe the impact of the research in this dissertation on the fields of developmental and evolutionary genetics and propose new lines of inquiry that could expand upon the findings within.

References

Abiola, Oduola, Joe M. Angel, Philip Avner, Alexander A. Bachmanov, John K. Belknap, Beth Bennett, Elizabeth P. Blankenhorn, et al. 2003. "The Nature and Identification of Quantitative Trait Loci: A Community's View." *Nature Reviews. Genetics* 4 (11): 911–16.

Aiello, D., K. Patel, and E. Lasagna. 2018. "The Myostatin Gene: An Overview of Mechanisms of Action and Its Relevance to Livestock Animals." *Animal Genetics* 49 (6): 505–19.

Ali, Ali, Rafet Al-Tobasei, Daniela Lourenco, Tim Leeds, Brett Kenney, and Mohamed Salem. 2019. "Genome-Wide Association Study Identifies Genomic Loci Affecting Filet Firmness and Protein Content in Rainbow Trout." *Frontiers in Genetics* 10 (May): 386.

Andersen, Svend Olav. 2010. "Insect Cuticular Sclerotization: A Review." *Insect Biochemistry and Molecular Biology* 40 (3): 166–78.

An, Xiaopeng, Yuxuan Song, Shuhai Bu, Haidong Ma, Kexin Gao, Jinxing Hou, Shan Wang, Zhang Lei, and Binyun Cao. 2016. "Association of Polymorphisms at the microRNA Binding Site of the Caprine KITLG 3'-UTR with Litter Size." *Scientific Reports* 6 (May): 25691.

Arif, Saad, Sebastian Kittelmann, and Alistair P. Mcgregor. 2015. "From Shavenbaby to the Naked Valley: Trichome Formation as a Model for Evolutionary Developmental Biology." *Evolution and Development* 17 (1): 120–26.

Arif, Saad, Sophie Murat, Isabel Almudi, Maria D. S. Nunes, Diane Bortolamiol-becet, Naomi S. Mcgregor, James M. S. Currie, et al. 2013. "Evolution of Mir-92a Underlies Natural Morphological Variation in *Drosophila Melanogaster*," 523–28.

Aust, Silvia, Florian Brüsselbach, Stefanie Pütz, and Bernhard T. Hovemann. 2010. "Alternative Tasks of *Drosophila* Tan in Neurotransmitter Recycling versus Cuticle Sclerotization Disclosed by Kinetic Properties." *The Journal of Biological Chemistry* 285 (27): 20740–47.

- Barrangou, Rodolphe, and Philippe Horvath. 2017. "A Decade of Discovery: CRISPR Functions and Applications." *Nature Microbiology* 2 (June): 17092.
- Bartel, David P. 2004. "MicroRNAs: Genomics, Biogenesis, Mechanism, and Function." *Cell* 116 (2): 281–97.
- Bartel, David P. 2018. "Metazoan MicroRNAs." *Cell* 173 (1): 20–51.
- Bier, Ethan, Melissa M. Harrison, Kate M. O'Connor-Giles, and Jill Wildonger. 2018. "Advances in Engineering the Fly Genome with the CRISPR-Cas System." *Genetics* 208 (1): 1–18.
- Biessmann, H. 1985. "Molecular Analysis of the Yellow Gene (y) Region of *Drosophila Melanogaster*." *Proceedings of the National Academy of Sciences of the United States of America* 82 (21): 7369–73.
- Borycz, Janusz, Jolanta a. Borycz, Mohammed Loubani, and Ian a. Meinertzhagen. 2002. "Tan and Ebony Genes Regulate a Novel Pathway for Transmitter Metabolism at Fly Photoreceptor Terminals." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 22 (24): 10549–57.
- Carroll, Sean B. 2005. "Evolution at Two Levels: On Genes and Form." *PLoS Biology* 3 (7): e245.
- Charlier, C., K. Segers, L. Karim, T. Shay, G. Gyapay, N. Cockett, and M. Georges. 2001. "The Callipyge Mutation Enhances the Expression of Coregulated Imprinted Genes in Cis without Affecting Their Imprinting Status." *Nature Genetics* 27 (4): 367–69.
- Chen, Congying, Chenlong Liu, Xinwei Xiong, Shaoming Fang, Hui Yang, Zhiyan Zhang, Jun Ren, Yuanmei Guo, and Lusheng Huang. 2018. "Copy Number Variation in the MSRB3 Gene Enlarges Porcine Ear Size through a Mechanism Involving miR-584-5p." *Genetics, Selection, Evolution: GSE* 50 (1): 72.
- Chen, Kevin, and Nikolaus Rajewsky. 2007. "The Evolution of Gene Regulation by Transcription Factors and microRNAs." *Nature Reviews. Genetics* 8 (2): 93–103.
- Chen, Ya Wen, Shilin Song, Ruifen Weng, Pushpa Verma, Jan Michael Kugler, Marita Buescher, Sigrid Rouam, and Stephenm Cohen. 2014. "Systematic Study of *Drosophila* MicroRNA Functions Using a Collection of Targeted Knockout Mutations." *Developmental Cell* 31 (6): 784–800.
- Chung, Henry, and Sean B. Carroll. 2015. "Wax, Sex and the Origin of Species: Dual Roles of Insect Cuticular Hydrocarbons in Adaptation and Mating." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 822–30.

Chung, Henry, David W. Loehlin, Héloïse D. Dufour, Kathy Vaccarro, Jocelyn G. Millar, and Sean B. Carroll. 2014. "A Single Gene Affects Both Ecological Divergence and Mate Choice in *Drosophila*." *Science* 343 (6175): 1148–51.

Clop, Alex, Fabienne Marcq, Haruko Takeda, Dimitri Pirottin, Xavier Tordoir, Bernard Bibé, Jacques Bouix, et al. 2006. "A Mutation Creating a Potential Illegitimate microRNA Target Site in the Myostatin Gene Affects Muscularity in Sheep." *Nature Genetics* 38 (7): 813–18.

Cockett, N. E., S. P. Jackson, T. L. Shay, F. Farnir, S. Berghmans, G. D. Snowder, D. M. Nielsen, and M. Georges. 1996. "Polar Overdominance at the Ovine Callipyge Locus." *Science* 273 (5272): 236–38.

Cockett, N. E., S. P. Jackson, T. L. Shay, D. Nielsen, S. S. Moore, M. R. Steele, W. Barendse, R. D. Green, and M. Georges. 1994. "Chromosomal Localization of the Callipyge Gene in Sheep (*Ovis Aries*) Using Bovine DNA Markers." *Proceedings of the National Academy of Sciences of the United States of America* 91 (8): 3019–23.

Cockett, N. E., S. P. Jackson, G. D. Snowder, T. L. Shay, S. Berghmans, J. E. Beever, C. Carpenter, and M. Georges. 1999. "The Callipyge Phenomenon: Evidence for Unusual Genetic Inheritance." *Journal of Animal Science* 77 Suppl 2: 221–27.

Cooley, Arielle M., Laura Shefner, Wesley N. McLaughlin, Emma E. Stewart, and Patricia J. Wittkopp. 2012. "The Ontogeny of Color: Developmental Origins of Divergent Pigmentation in *Drosophila Americana* and *D. Novamexicana*." *Evolution & Development* 14 (4): 317–25.

Courtier-Orgogozo, Virginie, Laurent Arnoult, Stéphane R. Prigent, Séverine Wiltgen, and Arnaud Martin. 2019. "Gephebase, a Database of Genotype-Phenotype Relationships for Natural and Domesticated Variation in Eukaryotes." *Nucleic Acids Research* 48 (D1): D696–703.

Davis, Erica, Charlotte Harken Jensen, Henrik Daa Schroder, Frédéric Farnir, Tracy Shay-Hadfield, Anette Kliem, Noelle Cockett, Michel Georges, and Carole Charlier. 2004. "Ectopic Expression of DLK1 Protein in Skeletal Muscle of Padumnal Heterozygotes Causes the Callipyge Phenotype." *Current Biology: CB* 14 (20): 1858–62.

Dembeck, Lauren M., Wen Huang, Michael M. Magwire, Faye Lawrence, Richard F. Lyman, and Trudy F. C. Mackay. 2015. "Genetic Architecture of Abdominal Pigmentation in *Drosophila Melanogaster*." *PLoS Genetics* 11 (5). <https://doi.org/10.1371/journal.pgen.1005163>.

Doudna, J. a., and E. Charpentier. 2014. "The New Frontier of Genome Engineering with CRISPR-Cas9." *Science* 346 (6213): 1258096–1258096.

Freedman, David H. 2010. "Why Scientific Studies Are so Often Wrong: The Streetlight Effect." *Discover Magazine* 26.
<https://www.sjsu.edu/people/fred.prochaska/courses/ScWk240/s3/Freedman-Week-15.pdf>.

Gao, Yun-Qian, Xin Chen, Pei Wang, Lei Lu, Wei Zhao, Chen Chen, Cai-Ping Chen, et al. 2015. "Regulation of DLK1 by the Maternally Expressed miR-379/miR-544 Cluster May Underlie Callipyge Polar Overdominance Inheritance." *Proceedings of the National Academy of Sciences of the United States of America* 112 (44): 13627–32.

Georges, M., A. Clop, F. Marcq, H. Takeda, D. Pirottin, S. Hiard, X. Tordoir, et al. 2006. "Polymorphic MicroRNA–Target Interactions: A Novel Source of Phenotypic Variation." *Cold Spring Harbor Symposia on Quantitative Biology* 71 (January): 343–50.

Gratz, Scott J., Alexander M. Cummings, Jennifer N. Nguyen, Danielle C. Hamm, Laura K. Donohue, Melissa M. Harrison, Jill Wildonger, and Kate M. O'Connor-Giles. 2013. "Genome Engineering of *Drosophila* with the CRISPR RNA-Guided Cas9 Nuclease." *Genetics* 194 (4): 1029–35.

Gratz, Scott J., Jill Wildonger, Melissa M. Harrison, and Kate M. O'Connor-Giles. 2013. "CRISPR/Cas9-Mediated Genome Engineering and the Promise of Designer Flies on Demand." *Fly* 7 (4): 249–55.

Hardy, Serge, Vincent Legagneux, Yann Audic, and Luc Paillard. 2010. "Reverse Genetics in Eukaryotes." *Biology of the Cell / under the Auspices of the European Cell Biology Organization* 102 (10): 561–80.

Hoekstra, H. E. 2006. "Genetics, Development and Evolution of Adaptive Pigmentation in Vertebrates." *Heredity* 97 (3): 222–34.

Ikegawa, Shiro. 2012. "A Short History of the Genome-Wide Association Study: Where We Were and Where We Are Going." *Genomics & Informatics* 10 (4): 220–25.

Jia, Xinzhen, Huiran Lin, Qinghua Nie, Xiquan Zhang, and Susan J. Lamont. 2016. "A Short Insertion Mutation Disrupts Genesis of miR-16 and Causes Increased Body Weight in Domesticated Chicken." *Scientific Reports* 6 (November): 36433.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (August): 816–22.

John, Alisha V., Lisa L. Sramkoski, Elizabeth A. Walker, Arielle M. Cooley, and Patricia J. Wittkopp. 2016. "Sensitivity of Allelic Divergence to Genomic Position: Lessons from the *Drosophila* Tan Gene." *G3* 6 (9): 2955–62.

- Kalay, Gizem, Richard Lusk, Mackenzie Dome, Korneel Hens, Bart Deplancke, and Patricia J. Wittkopp. 2016. "Potential Direct Regulators of the Drosophila Yellow Gene Identified by Yeast One-Hybrid and RNAi Screens." *G3* 6 (10): 3419–30.
- Kane, Nanci S., Mehul Vora, Krishna J. Varre, and Richard W. Padgett. 2017. "Efficient Screening of CRISPR/Cas9-Induced Events in Drosophila Using a Co-CRISPR Strategy." *G3* 7 (1): 87–93.
- Kennell, Jennifer a., Ken M. Cadigan, Iryna Shakhmantsir, and Evan J. Waldron. 2012. "The microRNA miR-8 Is a Positive Regulator of Pigmentation and Eclosion in Drosophila." *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 241 (1): 161–68.
- Kopp, A., I. Duncan, D. Godt, and S. B. Carroll. 2000. "Genetic Control and Evolution of Sexually Dimorphic Characters in Drosophila." *Nature* 408 (6812): 553–59.
- Kronforst, Marcus R., Gregory S. Barsh, Artyom Kopp, James Mallet, Antónia Monteiro, Sean P. Mullen, Meredith Protas, Erica B. Rosenblum, Christopher J. Schneider, and Hopi E. Hoekstra. 2012. "Unraveling the Thread of Nature's Tapestry: The Genetics of Diversity and Convergence in Animal Pigmentation." *Pigment Cell and Melanoma Research*. John Wiley & Sons, Ltd. <https://doi.org/10.1111/j.1755-148X.2012.01014.x>.
- Lamb, Abigail M., Elizabeth A. Walker, and Patricia J. Wittkopp. 2017. "Tools and Strategies for Scarless Allele Replacement in Drosophila Using CRISPR/Cas9." *Fly* 11 (1): 53–64.
- Lamb, Abigail M., Zinan Wang, Patricia Simmer, Henry Chung, and Patricia J. Wittkopp. 2020. "Ebony Affects Pigmentation Divergence and Cuticular Hydrocarbons in Drosophila Americana and D. Novamexicana." *Frontiers in Ecology and Evolution* 8: 184.
- Lucas, Keira, and Alexander S. Raikhel. 2013. "Insect microRNAs: Biogenesis, Expression Profiling and Biological Functions." *Insect Biochemistry and Molecular Biology* 43 (1): 24–38.
- Massey, J. H., and P. J. Wittkopp. 2016. "The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species." In *Current Topics in Developmental Biology*, 119:27–61.
- Massey, Jonathan H., Noriyoshi Akiyama, Tanja Bien, Klaus Dreisewerd, Patricia J. Wittkopp, Joanne Y. Yew, and Aya Takahashi. 2019. "Pleiotropic Effects of Ebony and Tan on Pigmentation and Cuticular Hydrocarbon Composition in Drosophila Melanogaster." *Frontiers in Physiology* 10 (MAY): 518.
- Miska, Eric A., Ezequiel Alvarez-Saavedra, Allison L. Abbott, Nelson C. Lau, Andrew B. Hellman, Shannon M. McGonagle, David P. Bartel, Victor R. Ambros, and H. Robert

Horvitz. 2007. "Most *Caenorhabditis Elegans* microRNAs Are Individually Not Essential for Development or Viability." *PLoS Genetics* 3 (12): e215.

Monteiro, Antónia. 2015. "Origin, Development, and Evolution of Butterfly Eyespots." *Annual Review of Entomology* 60 (January): 253–71.

Novitski, Edward, and M. Ashburner. 1976. *The Genetics and Biology of Drosophila*. London ; New York: Academic Press.

Paneru, Bam Dev, Rafet Al-Tobasei, Brett Kenney, Timothy D. Leeds, and Mohamed Salem. 2017. "RNA-Seq Reveals MicroRNA Expression Signature and Genetic Polymorphism Associated with Growth and Muscle Quality Traits in Rainbow Trout." *Scientific Reports* 7 (1): 9078.

Port, Phillip, Nadine Muschalik, and Simon L. Bullock. 2015. "Systematic Evaluation of *Drosophila* CRISPR Tools Reveals Safe and Robust Alternatives to Autonomous Gene Drives in Basic Research." *G3* 5 (7): 1493–1502.

Qiu, Mohan, Zengrong Zhang, Xia Xiong, Huarui Du, Qingyun Li, Chunlin Yu, Wu Gan, et al. 2020. "High-Throughput Sequencing Analysis Identified microRNAs Associated with Egg Production in Ducks Ovaries." *PeerJ* 8 (February): e8440.

Rebeiz, Mark, and Thomas M. Williams. 2017. "Using *Drosophila* Pigmentation Traits to Study the Mechanisms of Cis-Regulatory Evolution." *Current Opinion in Insect Science* 19: 1–7.

Rockman, Matthew V. 2012. "The QTN Program and the Alleles That Matter for Evolution: All That's Gold Does Not Glitter." *Evolution; International Journal of Organic Evolution* 66 (1): 1–17.

Rogers, William a., Sumant Grover, Samantha J. Stringer, Jennifer Parks, Mark Rebeiz, and Thomas M. Williams. 2014. "A Survey of the Trans-Regulatory Landscape for *Drosophila Melanogaster* Abdominal Pigmentation." *Developmental Biology* 385 (2): 417–32.

Schertel, Claus, Tobias Rutishauser, Klaus Förstemann, and Konrad Basler. 2012. "Functional Characterization of *Drosophila* microRNAs by a Novel in Vivo Library." *Genetics* 192 (4): 1543–52.

Sobel, James M., and Matthew A. Streisfeld. 2013. "Flower Color as a Model System for Studies of Plant Evo-Devo." *Frontiers in Plant Science* 4 (August): 321.

Stern, David L. 2014. "Identification of Loci That Cause Phenotypic Variation in Diverse Species with the Reciprocal Hemizyosity Test." *Trends in Genetics: TIG* 30 (12): 547–54.

Stern, David L., Justin Crocker, Yun Ding, Nicolas Frankel, Gretchen Kappes, Elizabeth Kim, Ryan Kuzmickas, Andrew Lemire, Joshua D. Mast, and Serge Picard. 2017. "Genetic and Transgenic Reagents for *Drosophila Simulans*, *D. Mauritiana*, *D. Yakuba*, *D. Santomea*, and *D. Virilis*." *G3* 7 (4): 1339–47.

Stern, David L., and Nicolás Frankel. 2013. "The Structure and Evolution of Cis-Regulatory Regions: The Shavenbaby Story." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 368 (1632): 20130028.

Stern, David L., and Virginie Orgogozo. 2008. "The Loci of Evolution: How Predictable Is Genetic Evolution?" *Evolution; International Journal of Organic Evolution* 62 (9): 2155–77.

Stern, D. L. 1998. "A Role of Ultrabithorax in Morphological Differences between *Drosophila* Species." *Nature* 396 (6710): 463–66.

Sturtevant, A. H. 1913. "The Linear Arrangement of Six Sex-Linked Factors in *Drosophila*, as Shown by Their Mode of Association." *The Journal of Experimental Zoology*, publ. 36, 14 (1): 43–59.

Sun, Dan, Zhaojiang Guo, Yong Liu, and Youjun Zhang. 2017. "Progress and Prospects of CRISPR/Cas Systems in Insects and Other Arthropods." *Frontiers in Physiology* 8 (September): 608.

Takahashi, Aya. 2013. "Pigmentation and Behavior: Potential Association through Pleiotropic Genes in *Drosophila*." *Genes & Genetic Systems* 88 (3): 165–74.

Takeda, Haruko, Florian Caiment, Maria Smit, Samuel Hiard, Xavier Tordoir, Noelle Cockett, Michel Georges, and Carole Charlier. 2006. "The Callipyge Mutation Enhances Bidirectional Long-Range DLK1-GTL2 Intergenic Transcription in Cis." *Proceedings of the National Academy of Sciences of the United States of America* 103 (21): 8119–24.

True, John R. 2003. "Insect Melanism: The Molecules Matter." *Trends in Ecology and Evolution*. Elsevier Ltd. <https://doi.org/10.1016/j.tree.2003.09.006>.

True, John R., Shu-Dan Yeh, Bernhard T. Hovemann, Tobias Kemme, Ian a. Meinertzhagen, Tara N. Edwards, Shian-Ren Liou, Qian Han, and Jianyong Li. 2005. "*Drosophila* Tan Encodes a Novel Hydrolase Required in Pigmentation and Vision." *PLoS Genetics* 1 (5): e63.

Walter, M. F., B. C. Black, G. Afshar, a. Y. Kermabon, T. R. Wright, and H. Biessmann. 1991. "Temporal and Spatial Expression of the Yellow Gene in Correlation with Cuticle Formation and Dopa Decarboxylase Activity in *Drosophila* Development." *Developmental Biology* 147 (1): 32–45.

Wittkopp, Patricia J., and Patrícia Beldade. 2009. "Development and Evolution of Insect Pigmentation: Genetic Mechanisms and the Potential Consequences of Pleiotropy." *Seminars in Cell and Developmental Biology*. Elsevier Ltd. <https://doi.org/10.1016/j.semcdb.2008.10.002>.

Wittkopp, Patricia J., Sean B. Carroll, and Artyom Kopp. 2003. "Evolution in Black and White: Genetic Control of Pigment Patterns in *Drosophila*." *Trends in Genetics: TIG* 19 (9): 495–504.

Wittkopp, Patricia J., Emma E. Stewart, Lisa L. Arnold, Adam H. Neidert, Belinda K. Haerum, Elizabeth M. Thompson, Saleh Akhras, Gabriel Smith-Winberry, and Laura Shefner. 2009. "Supporting Material: Intraspecific Polymorphism to Interspecific Divergence: Genetics of Pigmentation in *Drosophila*." *Science* 326 (5952): 540–44.

Wittkopp, Patricia J., John R. True, and Sean B. Carroll. 2002. "Reciprocal Functions of the *Drosophila* Yellow and Ebony Proteins in the Development and Evolution of Pigment Patterns." *Development* 129 (8): 1849–58.

Wittkopp, Patricia J., Barry L. Williams, Jayne E. Selegue, and Sean B. Carroll. 2003. "Drosophila Pigmentation Evolution: Divergent Genotypes Underlying Convergent Phenotypes." *Proceedings of the National Academy of Sciences of the United States of America* 100 (4): 1808–13.

Wright, Theodore R. F. 1987. "The Genetics of Biogenic Amine Metabolism, Sclerotization, and Melanization in *Drosophila Melanogaster*." *Advances in Genetics* 24 (C): 127–222.

Wright, Theodore R. F., Bruce C. Black, Clifton P. Bishop, J. Lawrence Marsh, Ellen S. Pentz, Ruth Steward, and Eileen Y. Wright. 1982. "The Genetics of Dopa Decarboxylase in *Drosophila Melanogaster*." *Molecular & General Genetics: MGG* 188 (1): 18–26.

Wu, N., U. Gaur, Q. Zhu, B. Chen, Z. Xu, X. Zhao, M. Yang, and D. Li. 2017. "Expressed microRNA Associated with High Rate of Egg Production in Chicken Ovarian Follicles." *Animal Genetics* 48 (2): 205–16.

Xie, Xiaohui, Jun Lu, E. J. Kulbokas, Todd R. Golub, Vamsi Mootha, Kerstin Lindblad-Toh, Eric S. Lander, and Manolis Kellis. 2005. "Systematic Discovery of Regulatory Motifs in Human Promoters and 3' UTRs by Comparison of Several Mammals." *Nature* 434 (7031): 338–45.

Zhao, Yong, Eva Samal, and Deepak Srivastava. 2005. "Serum Response Factor Regulates a Muscle-Specific microRNA That Targets Hand2 during Cardiogenesis." *Nature* 436 (7048): 214–20.

Chapter 2

Tools and Strategies for Scarless Allele Replacement in *Drosophila* Using CRISPR/Cas9ⁱ

Abstract

Genome editing via the CRISPR/Cas9 RNA-guided nuclease system has opened up exciting possibilities for genetic analysis. However, technical challenges associated with homology-directed repair have proven to be roadblocks for producing changes in the absence of unwanted, secondary mutations commonly known as “scars.” To address these issues, we developed a 2-stage, marker-assisted strategy to facilitate precise, “scarless” edits in *Drosophila* with a minimal requirement for molecular screening. Using this method, we modified 2 base pairs in a gene of interest without altering the final sequence of the CRISPR cut sites. We executed this 2-stage allele swap using a novel transformation marker that drives expression in the pupal wings, which can be screened for in the presence of common eye-expressing reporters. The tools we developed can be used to make a single change or a series of allelic substitutions in a region of interest in any *D. melanogaster* genetic background as well as in other *Drosophila* species.

ⁱ This chapter is published as: Lamb, Abigail M., Elizabeth A. Walker, and Patricia J. Wittkopp. 2017. “Tools and Strategies for Scarless Allele Replacement in *Drosophila* Using CRISPR/Cas9.” *Fly* 11 (1): 53–64.

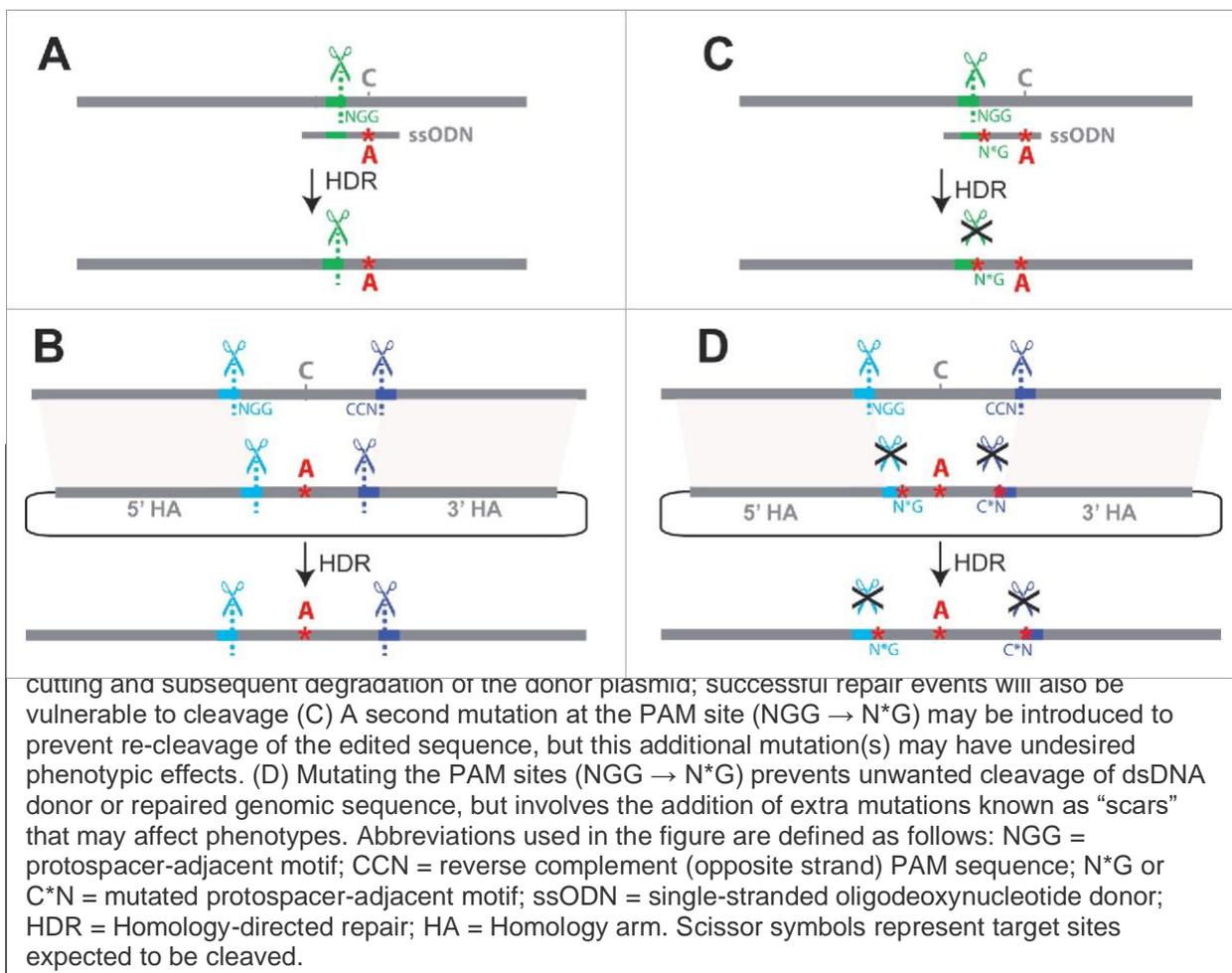
Introduction

In recent years, genome editing by site-specific nucleases has rapidly increased in accessibility, ease, and efficiency. Most notably, the CRISPR/Cas9 RNA-guided nuclease system has been developed and optimized for a variety of applications from basic gene disruption to knock-ins, endogenous tagging of proteins, and modulation of gene expression.¹⁻⁵ In the CRISPR/Cas9 system, adapted from *Streptococcus pyogenes* (henceforth shortened as “CRISPR”), the Cas9 endonuclease complexes with a single-guide RNA (sgRNA), which is designed by the experimenter to match a ~20bp target sequence, and causes double-strand cleavage 3 nucleotides 5' of the 3' end of the target sequence.⁵ The selection of sgRNA target sites is flexible, but a protospacer-adjacent motif (PAM) (the trinucleotide 5'-NGG-3') is required immediately 3' of the 20bp target sequence.⁵ After the double-strand break is induced at the target site, the cell's native DNA repair machinery can (i) rejoin the 2 ends of the break through non-homologous end joining (NHEJ), an error-prone process that frequently results in small insertions or deletions, or (ii) close the gap using homology-directed repair (HDR), in which a DNA molecule with sequence homologous to the break site is used as a repair template.³ The repair template used for double-strand break repair by HDR can be a homologous chromosome or an exogenous donor molecule with homology to both sides of the break site.

Exogenous repair templates, which are generally used to make one or more nucleotide changes in a sequence of interest, typically come in one of 2 forms: a single-stranded oligonucleotide donor (ssODN)¹ or a double-stranded DNA (dsDNA) plasmid.^{4,6} ssODNs are more quickly synthesized, but most synthesis companies limit their length to ~200bp, making them useful only when a suitable CRISPR target site is in close proximity to the site of interest (Figure 2-1A).^{3,4,6,7} dsDNA plasmids can be much larger than ssODNs, allowing modifications to be made at a greater distance from the cleavage site as well as allowing larger sequences to be inserted at the site of repair.^{4,6} To modify larger regions of sequence, 2 target sites can be used to cut out the region to be modified. In this case, the dsDNA used for repair contains the desired

change(s) flanked by homology arms targeting regions of DNA outside the 2 target sites (Figure 2-1B). Regardless of which type of repair template is used, the end result of homology-directed repair (HDR) is that the DNA sequence included in the repair template is incorporated into the native locus.

Using CRISPR-induced HDR to edit specific nucleotides is not always straightforward, however. A major challenge results from the fact that the target site sequence recognized by the sgRNA that directs the initial double-stranded break in the genome is also typically present in the repair template (a.k.a. donor DNA) as well as in the genome after editing. These sequences also interact with the sgRNA-Cas9 complex and experience unwanted cleavage, complicating the process of editing nucleotides in a single step of HDR (Figure 2-1A, B). This unwanted cleavage can be prevented by introducing one or more secondary changes that ablate the PAM site or alter the target site in the donor DNA in addition to the desired sequence edits, preventing Cas9 from



cleaving the donor as well as the genome after it has been edited (Figure 2-1C, D).^{4,8,9} Because these secondary mutations, or “scars,” remain in the genome after editing, care must be taken to minimize their phenotypic effects. When the desired sequence edits are located in (or near) a coding sequence, a synonymous change(s) can be introduced as the secondary change(s) to prevent recutting.^{4,9,10} Synonymous changes are often assumed to have little impact on protein function, but they do have phenotypic effects in some cases.^{11–13} When the desired changes are located in a non-coding region far from coding sequence, the PAM site can be ablated with a secondary non-coding change, but the impact of non-coding changes is even more difficult to predict.¹¹ Ideally, HDR should be used to change only the desired nucleotide(s) without introducing any other changes to the final modified genome (i.e. “scarless” editing).

A second challenge when using CRISPR to modify genomes is identifying individuals that have successfully inherited the desired genome alterations. This challenge is especially acute for modifications that require HDR because HDR resolution of double-strand breaks tends to be more rare than non-homologous end joining (NHEJ)¹⁴ and imprecise HDR often occurs that results in unwanted changes such as incorporating additional DNA at the edited locus.¹⁵ In multicellular species, only genome edits present in germ cells can be transmitted to offspring, and each individual carrying a desired genome modification in its germline (a “founder”) may transmit it to only a small percentage of its progeny.^{1,6,16} Molecular techniques such as high-resolution melt analysis, Surveyor assays, or even complete sequencing of the targeted region can be used to identify F₁ individuals with edited genomes and are especially useful in cultured cells that can be propagated throughout the screening process.^{4,6,17} Molecular screening methods can be costly and time consuming in a species such as *Drosophila*, however, where F₁ progeny are usually genotyped only after F₂ progeny are produced because the F₁ individuals must be killed to extract DNA and perform genotyping most reliably. As an alternative, *Drosophila* researchers often choose to incorporate a visible transformation marker for phenotypic screening.³

When gene deletion, disruption or tagging are desired, a selectable marker may be permanently integrated at the targeted locus, but if the goal is to determine the

effects of precise nucleotide changes, any transformation markers used should be removed prior to phenotypic analysis. Recombinase-mediated cassette exchange (RMCE) has been used to remove and replace such a marker gene in *Drosophila*¹⁸; however, RMCE is not ideal for this purpose because it leaves scars in the form of 2 attR sequences at the site of reporter excision. A related strategy has recently been developed in which the transformation marker is flanked by repeats that are recognized by the PiggyBac transposase (<http://flycrispr.molbio.wisc.edu/scarless>). PiggyBac-mediated excision leaves a TTAA motif behind, thus this method can be used to remove the transformation marker in a scarless manner when the locus of interest contains an endogenous TTAA motif. An alternative scarless “pop-in/pop-out” strategy was recently described for use in mammalian cells in which a fluorescent reporter gene is inserted along with the desired nucleotide changes in a first transformation step (pop-in), allowing modified cells to be identified using fluorescence-activated cell sorting (FACS), and then removed via a second round of CRISPR editing (pop-out).^{19,20}

Here, we describe an alternative pop-in/pop-out strategy for precisely editing one or more nucleotides in *Drosophila*. This method requires less molecular screening than single-stage allele replacement strategies and does not result in any unwanted sequence changes in the genome (i.e., it is scarless). It uses a customizable intermediate donor plasmid with a fluorescent reporter gene for easily identifying germline transmission of HDR events that is expressed in the pupal wings, allowing screening for it in the presence of the widely used eye-expressing fluorescent markers. This reporter gene is then cleanly replaced with the desired sequence, resulting in a “scarless” allele swap. We have successfully used this method to introduce 2 single-nucleotide changes into a *D. melanogaster* genome without inducing any additional modifications. All components used for these reactions (dsDNA repair templates and plasmids encoding Cas9 and sgRNAs) were co-injected into the embryo, making this method suitable for use in any strain of *D. melanogaster* as well as in other *Drosophila* species.

Results

The potential of using CRISPR to modify as few as one nucleotide in the genome makes it a powerful tool for testing the phenotypic consequences of changes in DNA sequence ranging from single nucleotide variants to more substantial differences in haplotype. We developed our strategy for making precise nucleotide changes in *Drosophila* by using CRISPR to modify a strain of *D. melanogaster* (*mel^{DA tan}*) that carries a *Drosophila americana* allele of the *tan* gene in a PiggyBac transgene marked with an eye-expressing green fluorescent protein (GFP). This strain also carries loss-of-function mutations in the endogenous *D. melanogaster* *yellow*, *white*, and *tan* genes (*ywt*), with the *D. americana tan* transgene rescuing the *tan* mutant phenotype. The *D. americana tan* allele was inserted into *D. melanogaster* to study changes in its first intron that contribute to pigmentation divergence between *D. americana* and its sister

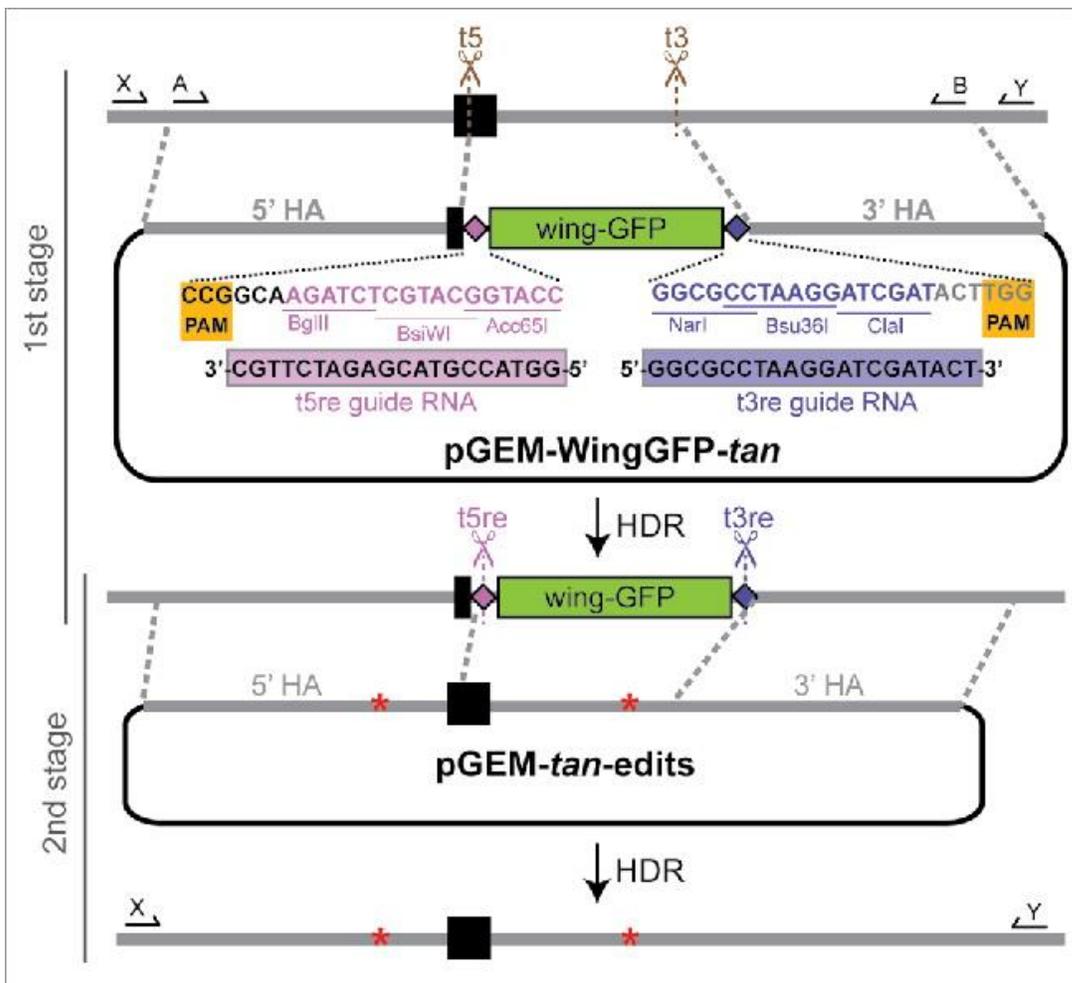


Figure 2-2: Schematic of marker-assisted, 2-stage allele swap within the *D. americana tan* transgene in *D. melanogaster* (previous page). In stage 1, the 3' end of the first exon (black rectangle) and a portion of the first intron of *D. americana tan* were excised by cleavage at the t5 (antisense direction) and t3 sgRNA target sites shown in brown. In the donor plasmid used to repair this region, pGEM-WingGFP-*tan*, the PAM sites (highlighted in yellow) and 3 PAM-proximal nucleotides at each target site contain sequence from the native target site, but the remaining 17 nucleotides of each sgRNA target sequence have been edited to differ from the *D. americana tan* sequence. These edited sequences serve as new, unique CRISPR target sites for reporter excision, and are labeled as t5re and t3re (“re” for “reporter excision”). These t5re and t3re target sites are not recognized by the sgRNA-Cas9 complexes targeting sites t5 and t3, thus preventing cleavage of pGEM-WingGFP-*tan* or the HDR product. When the wing-GFP transformation marker was incorporated into the genome, so were these unique t5re and t3re target sites, which contain restriction sites that double as multiple cloning regions. The donor plasmid used for stage 2, pGEM-*tan*-edits, contained the region of the *D. americana tan* sequence amplified with primers shown as arrows labeled A and B, which was cloned into the pGEM T-Easy vector. Changes in the length of 2 homopolymer runs used to confirm genome modification are represented by red asterisks. sgRNAs targeting the t5re and t3re sites flanking the reporter gene were used to remove it, with the *D. americana tan* sequence restored from pGEM-*tan*-edits via HDR. Locations of PCR primers used to test flies that lost wing-GFP expression following stage 2 of the allele swap are shown with arrows labeled X and Y. Precise HDR was confirmed by Sanger sequencing the amplicon produced by these primers. For primer sequences and details about screening PCRs, see Supplemental Table S2-1 in the supplement. Sanger sequencing chromatograms for all edited sites are shown in Supplemental Figure S2-1.

species *D. novamexicana*,²¹ thus we targeted the first intron of *tan* for genome modification when developing the tools described below.

First, we identified unique CRISPR target sites flanking our region of interest by using the flyCRISPR Optimal Target Finder tool to rule out target sites with sequence similarity elsewhere in the genome that might cause off-target cleavage.²² We chose to place the 5' target site within the first exon of the *tan* transgene so that the pigmentation phenotype could serve as a secondary indicator of successful gene disruption and later repair (Figure 2-2). The 3' target site was located in the first intron of *tan*. Each of these 2 selected target sequences was then cloned into its own sgRNA expression plasmid (pCFD3).²³

Next, we constructed an intermediate donor plasmid designed to replace our region of interest with a visible transformation marker flanked by unique CRISPR target sites (Figure 2-2). The *meI^{DA} tan* strain we sought to modify already contained an eye-expressing red fluorescent protein (RFP) reporter gene marking the attP landing-site used to insert the piggyBac transgene and an eye-expressing GFP marking the transgene, so we chose a transformation marker for CRISPR that expressed a fluorescent protein in a tissue other than the eye: a GFP reporter gene under the control

of a ~1kb enhancer of the *D. melanogaster yellow* gene that drives robust expression in pupal wings.²⁴ This reporter gene was amplified from genomic DNA of a previously constructed transgenic line using primers with sequences designed to function as unique CRISPR target sites appended at the 5' and 3' ends (Figure 2-2). These unique CRISPR target sites showed no exact matches (using a NCBI BLAST search) in any sequenced *Drosophila* genome, making them suitable for use in most, if not all, *Drosophila* species. They also contained cut sites for restriction enzymes, which can be used to easily remove and replace the wing-expressing GFP reporter gene with a different transformation marker.

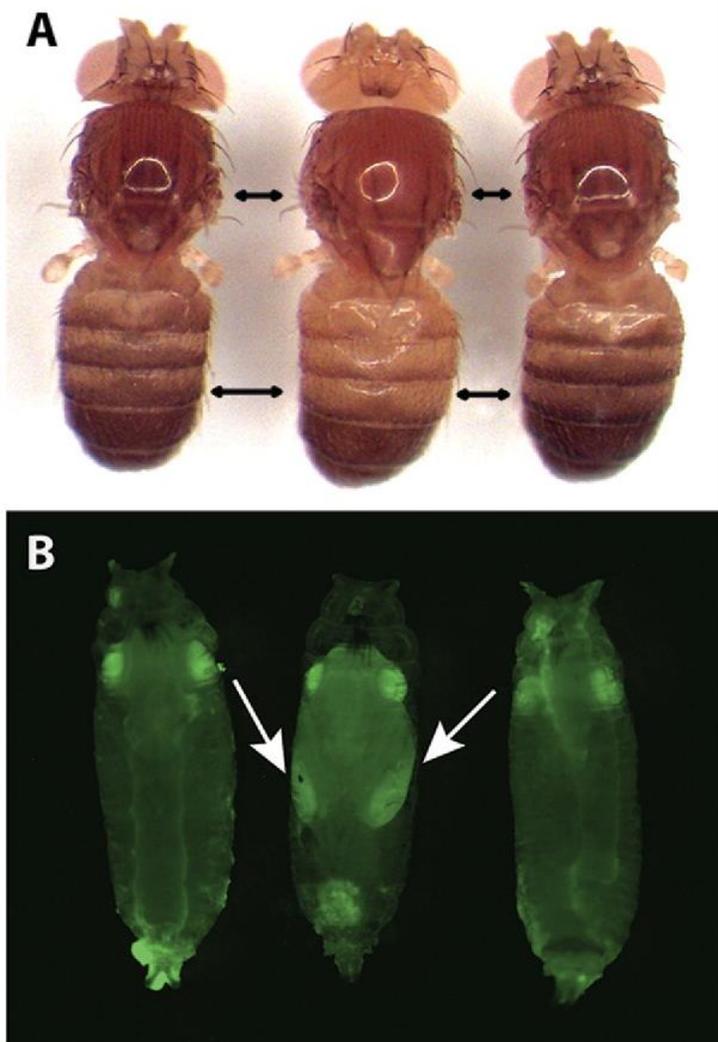
To direct this reporter gene to the desired region of the genome, homology arms adjacent to the original CRISPR target sites in the *tan* transgene were amplified and attached to the ends of this reporter gene flanked by unique CRISPR target sites in the pGEM T-Easy plasmid using Gibson Assembly (Figure 2-2).²⁵ The restriction enzyme cut sites in the pGEM T-Easy part of this pGEM-WingGFP-*tan* donor plasmid can be used in combination with the restriction enzyme cut sites in the unique CRISPR target sequences to easily replace one or both homology arms with different sequences for other studies (Figure 2-2).

We injected 1220 *meI^{DA tan}* embryos with a mixture of the pGEM-WingGFP-*tan* plasmid, both sgRNA expression plasmids targeting the *D. americana tan* transgene sequence, and a pBS-Hsp70-Cas9 expression plasmid producing Cas9 protein. We crossed 150 of the adult flies that emerged from these injected embryos back to the *meI^{DA tan}* strain and screened their F₁ progeny for inheritance of the reporter gene by looking for GFP expression in the developing pupal wings daily under a GFP-enabled stereoscope (Figure 2-3B). Six of these 150 injected flies produced progeny with GFP expression in pupal wings and were thus considered “founders.” The percentage of progeny expressing the WingGFP reporter construct from each founder ranged from 2.5% to 25.4%. In all, 70 pupae were positive for wing GFP expression, 43 of which were ultimately used to establish lines homozygous for the wing-expressing GFP marker gene. A summary of these statistics is provided in Supplemental Table S2-2.

All 43 of these homozygous lines showed lighter body pigmentation than the *D. mel^{DA tan}* parental line, consistent with the marker gene disrupting the *D. americana tan* transgene in an otherwise *tan* mutant *D. melanogaster* genetic background (Figure 2-3A). The pGEM T-Easy vector backbone was found to have been incorporated along with the marker gene in 28 of these 43 lines, and these lines were excluded from further study. PCR amplifications were then used to check the 5' and 3' insertion sites of the WingGFP marker in the remaining 15 lines, 13 of which were found to have incorporated it into the correct genomic location. DNA sequencing subsequently confirmed that the GFP reporter gene, unique CRISPR target sites, and homology regions were as expected in all 13 of these lines. Two of these 13 lines (25.17 and 9.14) were expanded for embryo collection and subsequent injection to excise the GFP reporter gene and replace it with a modified *tan* sequence.

A second donor plasmid was designed to restore the function of the *D. americana tan* transgene by replacing the wing-expressing GFP transformation marker with *tan* sequence excised in the first step. This plasmid, pGEM-*tan*-edits, was constructed by amplifying the *D. americana tan* transgene sequence from the beginning of the 5' homology arm upstream of exon 1 to the end of the 3' homology arm in the first intron (which includes the original CRISPR target sites) and cloning it into the pGEM T-Easy plasmid (Figure 2-2). The specific amplicon chosen to construct this plasmid contained 2 changes in non-coding homopolymer runs (9T->10T in the 5' homology arm and 10T->9T in the intron) that were not expected to affect the function of this sequence (Figure 2-2). These changes were included to allow us to confirm that the transformants recovered were not contaminants from the original *mel^{DA tan}* strain. Guide RNAs matching the 2 unique CRISPR target sites introduced with the GFP transformation marker were also cloned into the pCFD3 sgRNA expression plasmid.

We injected 631 embryos from the wing-GFP expressing line 25.17 and 730 from the wing-GFP expressing line 9.14 with a mixture containing the pBS-Hsp70-Cas9 expression plasmid, pCFD3 sgRNA expression plasmids targeting the unique CRISPR target sites, and pGEM-*tan*-edits. We crossed 179 flies emerging from these injected embryos to the same *ywt* strain of *D. melanogaster* that harbored the original *D.*



americana tan transgene.

Pupae from these crosses were screened for the absence of wing-expressing GFP along with the presence of eye-expressing GFP and RFP indicating presence of the *tan* transgene and the attP landing site the transgene was inserted into, respectively. From the 15 crosses found to contain one or more pupae that met these criteria, we collected a total of 32 pupae with this pattern of fluorescence before eclosion. Adult flies emerging from these pupae were crossed to a third chromosome (TM6B) balancer line and their progeny screened a second time for pupal wing

Figure 2-3: Representative pigmentation and fluorescence phenotypes of flies at each stage of the *tan* allele swap process (next page). (A) Dorsal pigmentation of adult flies is shown for the *mel^{DA tan}* strain prior to editing (left), the *mel^{DA tan}* strain in which the targeted region of *tan* has been replaced with the wing-GFP marker (middle), and the *mel^{DA tan}* strain after the wing-GFP marker was replaced with the edited *tan* sequence (right). The darker pigmentation seen in flies on the left and right is caused by a functional *D. americana tan* transgene. When this transgene is disrupted (middle), pigmentation is visibly lighter on the dorsal head, thorax, and abdomen. Double-headed black arrows indicate areas in the thorax and abdomen where the change in pigmentation was most readily apparent. (B) GFP fluorescence in late-stage pupae is shown for the *mel^{DA tan}* strain prior to editing (left), the *mel^{DA tan}* strain in which the targeted region of *tan* has been replaced with the wing-GFP marker (middle), and the *mel^{DA tan}* strain after the wing-GFP marker was replaced with the edited *tan* sequence (right). The GFP fluorescence in eyes of all 3 flies results from the 3XP3-GFP reporter gene included in the *D. americana tan* transgene. GFP expression in the developing wings (indicated with arrows) is visible in flies after the first stage of the 2-step allele swap procedure (middle) and lost following the second stage (right). All pupae shown were deemed to be at the same developmental stage based on visible features of wing development, expression of 3XP3-GFP, and lack of pigmentation on the developing wing.

GFP expression to make sure that the transient wing fluorescence was not simply missed during the initial screen. Ultimately, 5 founders produced 11 flies whose progeny were verified to have lost wing-expressing GFP. Ten of these 11 progeny were successfully used to establish lines homozygous for the edited transgene (Figure 2-3B), none of which showed evidence of the pGEM T-Easy backbone being incorporated. One of these 10 lines showed the ~2.5kb product expected from a PCR spanning the edited region from one homology region to the other and had dark pigmentation consistent with a rescue of *D. americana tan* function (Figures. 2-2, 2-3A). The remaining 9 lines failed to produce the expected PCR product and had light pigmentation suggesting that the *tan* sequence was not successfully restored. Sanger sequencing of the modified region of *D. americana tan* in the line with dark pigmentation showed precise repair with both homopolymer runs matching the donor plasmid sequence rather than the original transgene sequence (Supplemental Figure S2-1), confirming that we successfully introduced 2 single-nucleotide changes in the *D. americana tan* transgene in our desired *D. melanogaster* genetic background. A summary of efficiency at each stage of this second swap is provided in Supplemental Table S2-3. The fact that mutations in both the 5' homology arm and the intron were incorporated suggests that the HDR was likely initiated from a double-strand break at the t3re site. Because HDR is initiated by one of the two free 3' ends at the double-strand break, repair from the t5re site could only result in the incorporation of one or the other of these mutations, which are positioned to either side of the t5re site, whereas repair from the t3re site could incorporate both mutations.²⁶ We mention the directionality of repair to illustrate the importance of careful experimental design regarding the position of desired insertions/mutations relative to CRISPR target sites.

All of the reagents used in this work were developed with flexibility for future studies in mind. For example, restriction sites were included in the unique CRISPR target sequences for easy modifications, as described above. We have already used these restriction sites to make an alternative version of the intermediate donor vector (pGEM-WingGFP-*tan*) in which the pupal wing-expressing GFP reporter gene was replaced with an eye-expressing RFP (pGEM-3XP3.RFP-*tan*). Fluorescent proteins expressed in the adult eye by the 3XP3 promoter have been shown to function in a wide

variety of insect taxa,^{27,28} making this donor plasmid useful for HDR not only in *D. melanogaster*, but also in many other insect species. The wing-GFP marker we used allows screening in lines that already carry eye-expressing markers or have eye color that makes the detection of eye-expressing fluorescent markers difficult, but we encourage the use of the 3xP3-RFP intermediate donor when screening for fluorescence in the eyes of *white* mutant flies is possible because it is less laborious than screening for the gain and loss of expression from the pupal wing-GFP marker. pGEM-3XP3.RFP-*tan* also contains restriction sites in novel CRISPR target sites introduced to prevent re-cutting and facilitate easy cloning. The homology arms in this plasmid target the *D. americana tan* gene, but other researchers can replace these homology arms with their own sequences of interest. When preparing reagents for a new locus, it should be noted that the 3 PAM-proximal nucleotides of these unique CRISPR target sequences are specific to each locus, and the sgRNAs should be customized to match the locus targeted by the donor plasmid (Figure 2-2).

Discussion

We have developed tools and protocols to implement a 2-step, marker-assisted genome editing strategy suitable for making precise changes at targeted sites in *Drosophila* with greatly reduced requirements for molecular screening (Figure 2-4). Our method adds to the few available techniques that leave no unwanted changes (“scars”) in the genome, such as those that occur when ablating PAM sites or using integrase-mediated excision to remove selectable markers (<http://flycrispr.molbio.wisc.edu/scarless>).^{19,20} Our method is also better-suited than these other methods for making a series of allelic changes at the same locus, as the intermediate line containing the marker gene need only be generated once. This is useful, for example, when reintroducing the original, unedited sequence in parallel with an experimental manipulation as a control for side effects of the CRISPR process or when testing a set of allelic variants to identify sites with specific functions.

With three scarless, 2-stage allele swap methods now described (Xi et al.,²⁰ <http://flycrispr.molbio.wisc.edu/scarless>, and this study), researchers should consider the differences among these methods when designing their own experiments. First, both our method and the pHD-ScarlessDsRed method (<http://flycrispr.molbio.wisc.edu/scarless>) are specifically optimized for use in *Drosophila*, whereas the “pop-in/pop-out” strategy described by Xi et al.²⁰ uses reagents designed for use in mammalian cells. Second, the pHD-ScarlessDsRed method requires the presence of a TTAA motif at the target locus for scarless editing, which adds some restriction to target site selection, but circumvents the need for a second round of injections when working with *D. melanogaster* because flies carrying the reporter can be crossed to existing transgenic lines that express the PiggyBac transposase.²⁹ However, to apply this method in species other than *D. melanogaster*, a

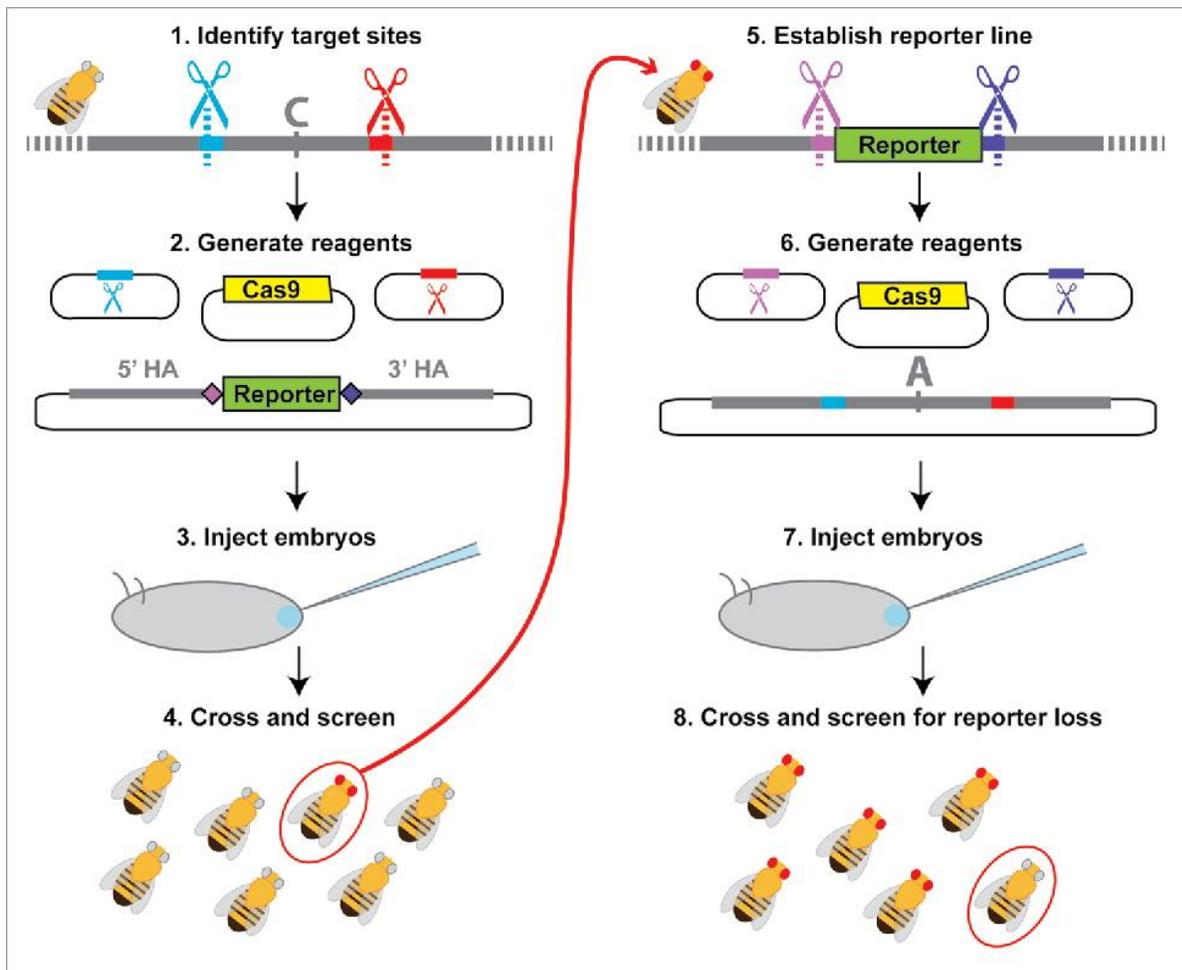


Figure 2-4: Workflow for 2-stage marker assisted allele swap. (1) Target sites flanking the area to be edited are identified (red and blue) using online tools to identify optimal target sites and search for potential off-target cleavage sites.^{22,39} (2) Sequences from the selected target sites are then cloned into sgRNA expression plasmids or used to generate *in vitro* transcribed sgRNAs. Homology arms flanking the region of interest (recommended length ~1kb) are cloned into the reporter donor plasmid, which contains unique CRISPR target sites (light and dark purple) in place of the genomic target sites (red and blue). (3) Embryos are injected with the donor plasmid, sgRNAs (expression plasmids or *in vitro* transcribed RNAs) and a source of Cas9 (expression plasmid, mRNA, or protein) unless any of these components is produced by a transgene already present in the host. (4) Adult flies that develop from the injected embryos are collected as virgins and then crossed back to the parental line. F₁ progeny emerging from these crosses are screened for the selectable marker, with flies positive for the selectable marker allowed to produce F₂ progeny before extracting their DNA for molecular screening. (5) Individuals with the correct reporter gene insertion are made homozygous and the population is expanded for embryo collections. (6) sgRNAs with sequences matching the unique CRISPR target sites introduced with the reporter gene (light and dark purple) as well as a plasmid containing the original CRISPR target sites and the edited version of the original sequence are prepared. (7) Flies carrying the reporter at the locus of interest are injected for the second allele swap step. (8) Because the selectable marker is dominant, adult flies developing from injected embryos must be crossed back to either the original parental line from step (1) or to a balancer line (if available) for screening. Progeny from this cross that do not show expression of the selectable marker are then crossed and analyzed with molecular tests to determine whether they contain the desired editing events.

second injection step will still be required in order to introduce the PiggyBac transposase. Third, our method uses a novel transformation marker and uniquely designed target sites in the reporter construct which double as cloning sites for later customization of homology arms or reporter cassette. Finally, with both the Xi et al.²⁰ and pHD-ScarlessDsRed methods, the desired changes are introduced along with a reporter gene during the first step and the reporter gene is excised in the second step. In our method, only the reporter gene is introduced in the first step. In instances where larger genomic regions are being edited, this feature reduces the size of the region that must be inserted initially by HDR, which increases efficiency,³⁰ and creates a stable genotype that can be used to eliminate the first stage CRISPR modification in any future experiments that modify the same locus.

Our method is also particularly well-suited for use in any *D. melanogaster* genetic background or in any *Drosophila* species. This feature realizes the great potential of genome editing via site-specific nucleases for making genetic manipulations at a gene's native locus and in its native genomic background. For example, the function of sites that have diverged between 2 *Drosophila* species can now be tested in their native context rather than in a heterologous species such as *D. melanogaster*.^{31–33} However,

currently available Cas9- and sgRNA-expression plasmids (including those used in this study) contain promoters derived from *D. melanogaster*, so injecting purified Cas9 protein or mRNA along with *in vitro* transcribed sgRNAs instead of using expression plasmids will likely give better results when working with other *Drosophila* species.^{17,34,35} We have recently had success using commercially available Cas9 protein and *in vitro* transcribed sgRNAs to induce NHEJ and/or HDR in *Drosophila elegans*, *Drosophila americana*, *Drosophila novamexicana*, and *Drosophila virilis* (unpublished data).

The two-stage allele swap method reported here provides additional precision and flexibility for allele replacements using CRISPR in *Drosophila*. Further modifications are likely to increase the efficiency of this method even more, however. For example, if a specific genetic background is not required, one of several lines of *D. melanogaster* developed to increase CRISPR efficiency can be used, such as lines with Cas9 and/or sgRNA expressed from transgenes integrated into the genome^{16,36,37} or lines with reduced *lig4* activity that increase the frequency of HDR events by inhibiting the NHEJ pathway.^{10,38} Similar lines could also be constructed in other *Drosophila* species to optimize CRISPR-based genome modifications in these hosts. Selection of sgRNA target sites may also be optimized to maximize the likelihood of cleavage according to criteria that have been identified in other studies.^{39–41} We note, however, that the need to improve CRISPR efficiency is decreased by the use of methods which employ fast and easy phenotypic screening of large populations.

In designing genome editing experiments, the most pertinent strategies and screening techniques will depend on the types of changes desired as well as the resources available to the researcher. For instance, experiments to alter the coding sequence of essential genes would rule out the use of an intermediate stage that disrupts both copies of the gene, which would preclude the use of our 2-stage method as described. Nonetheless, the applicability of our method to many other types of experiments in a wide variety of genetic backgrounds makes it a valuable addition to the existing methods and tools for scarless genome modifications available to the *Drosophila* research community.

Materials and methods

Fly strains

The *D. americana tan* transgene was constructed as previously described in Wittkopp et al.²¹ The transgenic strain of *D. melanogaster mel^{DA tan}* was constructed by integrating this *D. americana tan* transgene and a 3XP3-GFP transformation marker in a piggyBac plasmid containing an attB sequence into an attP landing site marked with 3XP3-RFP (Flybase ID FBst0024749) at cytological location 86Fb on the third chromosome using phi-C31-mediated integration.⁴² GenetiVision (Houston, TX) performed the injections that produced the *mel^{DA tan}* transgenic line. Transformant flies carrying the *mel^{DA tan}* transgene were crossed to a line that was mutant for *yellow*, *white*, and *tan* (*ywt*) to confirm that the *D. americana tan* transgene rescued the *D. melanogaster tan* mutant phenotype and to allow easier detection of the 3xP3-GFP and 3xP3-RFP fluorescent markers.

The transgenic *D. melanogaster* line carrying the wing-expressing GFP reporter gene (referred to as “line 890”) was constructed using *D. melanogaster yellow* enhancer sub-element “mel_a2” described in Kalay, 2012.²⁴ The reporter gene from line 890 was chosen as a selectable marker for this study because of its clear expression pattern in the developing wings, which can be screened independently of the eye-expressing fluorescent markers present in *mel^{DA tan}*.

To generate lines homozygous for the edited *D. americana tan* transgene on the third chromosomes, we first constructed a *ywt*;*+*;*TM6B* strain by crossing the *TM6B* third chromosome balancer (Flybase ID FBst0007197) into the same *ywt* genetic background used in the construction of *mel^{DA tan}*. This *ywt*;*+*;*TM6B* genotype was crossed with the originally recovered *mel^{DA tan}* flies to produce a stock homozygous for *ywt* as well as *mel^{DA tan}*. Flies were maintained at ~25°C on standard cornmeal media except were otherwise specified.

Plasmids

To construct pGEM-WingGFP-*tan*, the wing-expressing GFP reporter gene sequence from line 890 and homology arms flanking the targeted region of *D. americana tan* were PCR amplified to generate overlapping regions of homology for cloning into the pGEM T-easy vector via Gibson Assembly.²⁵ The 831bp 5' homology arm was PCR amplified from *D. americana tan* using primer pair 5 (Supplemental Table S2-1), which appended a region homologous to the pGEM T-easy vector on the 5' end and added new sequence to form the t5re target site on the 3' end (Figure 2-2, in light purple). The 966bp 3' homology arm was amplified using primer pair 6 (Supplemental Table S2-1), appending new sequence to form the t3re target site at the 5' end (Figure 2-2, in dark purple) and a region homologous to the pGEM T-easy vector to the 5' end. Both homology arm PCR reactions used *me^{DA tan}* genomic DNA as template. The wing-expressing GFP reporter was PCR amplified from line 890 genomic DNA using primer pair 7 (Supplemental Table S2-1), which appended the t5re target site to the 5' end and the t3re target site to the 3' end (Figure 2-2). These amplicons and the pGEM T-easy vector were assembled using New England Biolabs (NEB) Gibson Assembly Master Mix.

pGEM-*tan*-edits, the donor plasmid used for the second stage of the allele swap, was generated by PCR amplifying the targeted *D. americana tan* region along with the flanking homology regions from *me^{DA tan}* genomic DNA using primer pair 8 (Supplemental Table S2-1, see Figure 2-2) and inserting the resulting amplicon into pGEM T-Easy vector via Gibson Assembly.

To construct the eye-expressing RFP donor, pGEM-3XP3.RFP-*tan*, the 3XP3-RFP reporter was PCR amplified from *D. melanogaster* genomic DNA containing the M{3XP3-RFP.attP}ZH-51C landing site (Flybase ID FBtp0023088) using primer pair 11, which added Acc65I and Bsu36I restriction sites (Supplemental Table S2-1). Before constructing pGEM-3XP3.RFP-*tan*, we had replaced the 3' homology arm of pGEM-wingGFP-*tan* with sequence from another region of *D. americana tan* using Bsu36I and MluI restriction sites. This new homology arm was amplified from *me^{DA tan}* genomic DNA

using primer pair 10 (Supplemental Table S2-1). The 3XP3-RFP reporter amplicon was cloned into this modified pGEM-WingGFP-*tan* plasmid using the Acc65I and Bsu36I restriction enzyme cut sites.

sgRNA expression plasmids were made by ligating target-site specific annealed oligonucleotide inserts into BbsI-digested pCFD3 (Addgene # 49410) according to the methods described by Port et al.²³ The following oligonucleotide pairs were used to generate the cloning inserts for the indicated sgRNA target sites: 't5' – primer pair 12, 't3' – primer pair 13, 't5re' – primer pair 14, 't3re' – primer pair 15 (Supplemental Table S2-1). We used the pBS-Hsp70-Cas9 plasmid (Addgene #46294) as a Cas9 source.

Drosophila husbandry and injection

Plasmids for CRISPR were prepared for injection using either Zymo Zypzy Plasmid Maxi Prep kit or Mechery-Nagle Nucleobond Xtra EF Midi Prep kit followed by ethanol precipitation and re-suspension in nuclease-free water. For all injections, plasmid concentrations were as follows: 500ng/μL HDR donor, 100ng/μL each sgRNA plasmid, 250ng/μL pBS-Hsp70-Cas9. After injection, embryos were maintained at 25°C for 3–4 d, at which time larvae were moved to vials with cornmeal media. Embryo injections were performed as described previously.⁴³ For pupal wing reporter screening, flies were moved to 18°C upon entering the wandering larval stage to slow development in an effort to prolong the amount of time the fluorescent marker signal was present.

Fluorescence screening

To screen for the presence of fluorescent markers, we used a Leica MZ6 stereoscope equipped with a Kramer Scientific Quad Fluorescence Illuminator. GFP expression from the wing-GFP reporter gene used in this study becomes easily detectable in the wings after the developing wing is clearly visible, but before 3XP3-GFP signal is visible in the eyes. GFP signal in wings is easily detectable for approximately 2 d at 18°C, with GFP signal fading rapidly at the onset of wing pigmentation.

To screen for the presence or absence of the wing-expressing GFP marker, F₁ pupae (progeny of injected parents) were observed daily under the GFP stereoscope at 18°C. After the first stage of the allele swap, when the wing-GFP marker was inserted, pupae with detectable GFP expression in the developing wings were removed from the vial with a wet paintbrush and isolated in a ventilated microcentrifuge tube with food to await future crossing and molecular screening. Surviving wing-GFP positive pupae were crossed to the *ywt TM6B* balancer line. From these balancer crosses, siblings with both the wing-GFP phenotype and the TM6B bristle phenotype were crossed to form homozygous lines.

Following the second stage of the allele swap (marker excision and replacement), pupae with detectable GFP expression in the developing wings were removed and discarded. Any pupae that remained in the vial until their wings darkened were removed and isolated for future crossing. Crosses were performed as described in the results.

Molecular screening

To test for the presence of unwanted pGEM T-easy vector (plasmid “backbone”) in edited flies, we used PCR reactions with one primer in the vector backbone and the other in either the 5′ or 3′ homology arm, using primer pair 1 and primer pair 2 for the 5′ and 3′ sides, respectively (Supplemental Table S2-1), while the donor plasmid was used as a positive control template. Strains that produced a band from either of these PCR reactions were excluded from further study.

To confirm integration of the wing-GFP reporter gene into the correct genomic location, we used a PCR reaction that amplifies DNA sequence from within the reporter gene sequence to outside the homology region on both the 5′ and 3′ sides of the reporter gene, using primer pair 3 to screen the 5′ side and primer pair 4 to screen the 3′ side (Supplemental Table S2-1). The amplicons from these PCR reactions were Sanger sequenced to confirm scarless repair at both the target sites and throughout both homology regions. To screen for correct HDR after the second stage of the allele swap, the entire edited locus was amplified via PCR using primers outside the homology

regions (see Figure 2-2 and primer pair 9 in Supplemental Table S2-1 for details). This amplicon was Sanger sequenced to confirm the presence of expected sequence edits.

All diagnostic PCRs were performed using genomic DNA extracted from single flies following the Gloor and Engels “squish prep” protocol.⁴⁴

Imaging

Fly images shown in Figure 2-3 were captured using a Leica MZFLIII fluorescence stereoscope equipped with a Leica DC480 microscope camera.

Acknowledgments

We would like to thank Kathy Vaccarro and the Carroll Lab at University of Wisconsin for training and assistance in microinjection techniques, Scott Barolo and Lisa Johnson for assistance optimizing microinjection procedures, and Fabien Duveau, Andrea Hodgins-Davis, Alisha John, Jennifer Lachoweic, and Jonathan Massey for feedback and comments on this manuscript. This work was funded by the National Institute of Health (5-R01-GM-089736, P.J.W.), the National Institute of Health Genetics Training Grant (T32GM00754, A.M.L.), and the National Science Foundation Graduate Research Fellowship Program (DGC 1256260 A.M.L.).

Abbreviations

CRISPR	Conserved Regularly Interspaced Short Palindromic Repeats
Cas9	CRISPR-associated protein-9 nuclease

sgRNA	single guide RNA
PAM	protospacer-adjacent motif
ssODN	single-stranded oligodeoxynucleotide
dsDNA	double stranded DNA
NHEJ	Non-homologous end-joining
HDR	Homology-directed repair
GFP	green fluorescent protein
RFP	red fluorescent protein
SNP	single nucleotide polymorphism

References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; 337:816-22; PMID:22745249; <http://dx.doi.org/10.1126/science.1225829>
2. Bassett AR, Liu JL. CRISPR/Cas9 and Genome Editing in *Drosophila*. *J Genet Genomics* 2014; 41:7-19; PMID:24480743; <http://dx.doi.org/10.1016/j.jgg.2013.12.004>
3. Gratz SJ, Cummings AM, Nguyen JN, Hamm DC, Donohue LK, Harrison MM, Wi Idonger J, O'Connor-Giles KM. Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics* 2013; 194:1029-35; PMID:23709638; <http://dx.doi.org/10.1534/genetics.113.152710>
4. Ghorbal M, Gorman M, Macpherson CR, Martins RM, Scherf A, Lopez-Rubio JJ. Genome editing in the human malaria parasite *Plasmodium falciparum* using the CRISPR-Cas9 system. *Nat Biotechnol* 2014; 32:819-21; PMID:24880488; <http://dx.doi.org/10.1038/nbt.2925>
5. Inui M, Miyado M, Igarashi M, Tamano M, Kubo A, Yamashita S, Asahara H, Fukami M, Takada S. Rapid generation of mouse models with defined point mutations by the CRISPR/Cas9 system. *Sci Rep* 2014; 4:5396; PMID:24953798; <http://dx.doi.org/10.1038/srep05396>
6. Böttcher R, Hollmann M, Merk K, Nitschko V, Obermaier C, Philippou-Massier J, Wieland I, Gaul U, Förstemann K. Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. *Nucleic Acids Res* 2014; 42:e89; PMID:24748663; <http://dx.doi.org/10.1093/nar/gku289>
7. Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? *Evolution* 2008; 62:2155-77; PMID:18616572; <http://dx.doi.org/10.1111/j.1558-5646.2008.00450.x>
8. Stam LF, Laurie CC. Molecular dissection of a major gene effect on a quantitative trait: The level of alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics* 1996; 144:1559-64; PMID:8978044

9. Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskiy O, Makarov SS, Maixner W, Diatchenko L. Human Catechol-O-Methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 2006; 314:1930-3; PMID:17185601; <http://dx.doi.org/10.1126/science.1131262>
10. Miyaoka Y, Berman JR, Cooper SB, Mayerl SJ, Chan AH, Zhang B, Karlin-Neumann GA, Conklin BR. Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* 2016; 6:23549; PMID:27030102; <http://dx.doi.org/10.1038/srep23549>
11. Yu Z, Chen H, Liu J, Zhang H, Yan Y, Zhu N, Guo Y, Yang B, Chang Y, Dai F, et al. Various applications of TALEN- and CRISPR/Cas9-mediated homologous recombination to modify the *Drosophila* genome. *Biol Open* 2014; 3:271-80; PMID:24659249; <http://dx.doi.org/10.1242/bio.20147682>
12. Sebo ZL, Lee HB, Peng Y, Guo Y. A simplified and efficient germline-specific CRISPR/Cas9 system for *Drosophila* genomic engineering. *Fly (Austin)* 2014; 8:52-7; PMID:24141137; <http://dx.doi.org/10.4161/fly.26828>
13. Bassett AR, Tibbit C, Ponting CP, Liu J-L. Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. *Cell Rep* 2013; 4:220-8; PMID:23827738; <http://dx.doi.org/10.1016/j.celrep.2013.06.020>
14. Zhang X, Koolhaas WH, Schnorrer F. A versatile two-step CRISPR- and RMCE-based strategy for efficient genome engineering in *Drosophila*. *G3 (Bethesda)* 2014; 4:2409-18; PMID:25324299; http://dx.doi.org/full_text
15. Kühn R, Chu VT. Pop in, pop out: a novel gene-targeting strategy for use with CRISPR-Cas9. *Genome Biol* 2015; 16:244; PMID:26553112; <http://dx.doi.org/10.1186/s13059-015-0810-2>
16. Xi L, Schmidt JC, Zaug AJ, Ascarrunz DR, Cech TR. A novel two-step genome editing strategy with CRISPR-Cas9 provides new insights into telomerase action and TERT gene expression. *Genome Biol* 2015; 16:231; PMID:26553065; <http://dx.doi.org/10.1186/s13059-015-0791-1>
17. Wittkopp PJ, Stewart EE, Arnold LL, Neidert AH, Haerum BK, Thompson EM, Akhras S, Smith-Winberry G, Shefner L. Intraspecific polymorphism to interspecific

- divergence: Genetics of pigmentation in drosophila. *Science* 2009; 326:540-4; PMID:19900891; <http://dx.doi.org/10.1126/science.1176980>
18. Gratz SJ, Ukken FP, Rubinstein CD, Thiede G, Donohue LK, Cummings AM, O'Connor-Giles KM. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in *Drosophila*. *Genetics* 2014; 196:961-71; PMID:24478335; <http://dx.doi.org/10.1534/genetics.113.160713>
 19. Port F, Chen HM, Lee T, Bullock SL. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci U S A* 2014; 111:E2967-76; PMID:25002478; <http://dx.doi.org/10.1073/pnas.1405500111>
 20. Kalay G. Rapid evolution of cis-regulatory architecture and activity in the *Drosophila* yellow gene. [dissertation]. [Ann Arbor]: University of Michigan; 2012. 197 p.
 21. Gibson DG, Smith HO, Hutchison CA, Venter JC, Merryman C. Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* 2010; 7:901-3; PMID:20935651; <http://dx.doi.org/10.1038/nmeth.1515>
 22. Pardo B, Gómez-González B, Aguilera A. DNA repair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cell Mol Life Sci* 2009; 66:1039-56; PMID:19153654; <http://dx.doi.org/10.1007/s00018-009-8740-3>
 23. Holtzman S, Miller D, Eisman RC, Kuwayama H, Niimi T, Kaufman TC. Transgenic tools for members of the genus *Drosophila* with sequenced genomes. *Fly (Austin)* 2010; 4:349-62; PMID:20890112; <http://dx.doi.org/10.4161/fly.4.4.13304>
 24. Dong S, Lin J, Held NL, Clem RJ, Passarelli AL, Franz AW. Heritable CRISPR/Cas9-Mediated genome editing in the yellow fever mosquito, *Aedes aegypti*. *PLoS One* 2015; 10:e0122353; PMID:25815482; <http://dx.doi.org/10.1371/journal.pone.0122353>
 25. Griswold CM, Roebuck J, Anderson RO, Stam LF, Spana EP. A toolkit for transformation and mutagenesis in *Drosophila* using piggyBac. *Drosoph Inf Serv* 2002; 85:129-32

- 26.** Perez C, Guyot V, Cabaniols J-
P, Gouble A, Micheaux B, Smith J, Leduc S, Pâques F, Duchateau P. Factors affecting double-strand break-induced homologous recombination in mammalian cells. *Biotechniques* 2005; 39:109-15; PMID:16060375;
<http://dx.doi.org/10.2144/05391GT01>
- 27.** Wittkopp PJ, Vaccaro K, Carroll SB. Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr Biol* 2002; 12:1547-56; PMID:12372246;
[http://dx.doi.org/10.1016/S0960-9822\(02\)01113-2](http://dx.doi.org/10.1016/S0960-9822(02)01113-2)
- 28.** Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* 2008; 132:783-93; PMID:18329365;
<http://dx.doi.org/10.1016/j.cell.2008.01.014>
- 29.** Yassin A, Bastide H, Chung H, Veuille M, David JR, Pool JE. Ancient balancing selection at tan underlies female colour dimorphism in *Drosophila erecta*. *Nat Commun* 2016; 7:10400; PMID:26778363;
<http://dx.doi.org/10.1038/ncomms10400>
- 30.** Martin A, Serano JM, Jarvis E, Bruce HS, Wang J, Ray S, Barker CA, O'Connell LC, Patel NH. CRISPR/Cas9 mutagenesis reveals versatile roles of hox genes in crustacean limb specification and evolution. *Curr Biol* 2016; 26:14-26; PMID:26687626; <http://dx.doi.org/10.1016/j.cub.2015.11.021>
- 31.** Kistler KE, Vosshall LB, Matthews BJ. Genome engineering with CRISPR-Cas9 in the mosquito *Aedes aegypti*. *Cell Rep* 2015; 11:51-60; PMID:25818303;
<http://dx.doi.org/10.1016/j.celrep.2015.03.009>
- 32.** Port F, Muschalik N, Bullock SL. Systematic evaluation of *drosophila* CRISPR tools reveals safe and robust alternatives to autonomous gene drives in basic research. *G3 (Bethesda)* 2015; 5:1493-502; PMID:25999583;
http://dx.doi.org/full_text
- 33.** Ren X, Sun J, Housden BE, Hu Y, Roesel C, Lin S, Liu L-
P, Yang Z, Mao D, Sun L, et al. Optimized gene editing technology for *Drosophila melanogaster* using germ line-specific Cas9. *Proc Natl Acad Sci* 2013; 110:19012-7; PMID:24191015; <http://dx.doi.org/10.1073/pnas.1318481110>

- 34.** Beumer KJ, Trautman JK, Bozas A, Liu J-L, Rutter J, Gall JG, Carroll D. Efficient gene targeting in *Drosophila* by direct embryo injection with zinc-finger nucleases. *Proc Natl Acad Sci U S A* 2008; 105:19821-6; PMID:19064913; <http://dx.doi.org/10.1073/pnas.0810475105>
- 35.** Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* 2015; 16:218; PMID:26521937; <http://dx.doi.org/10.1186/s13059-015-0784-0>
- 36.** Ren X, Yang Z, Xu J, Sun J, Mao D, Hu Y, Yang S-J, Qiao H-H, Wang X, Hu Q, et al. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in *Drosophila*. *Cell Rep* 2014; 9:1151-62; PMID:25437567; <http://dx.doi.org/10.1016/j.celrep.2014.09.044>
- 37.** Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 2014; 32:1262-7; PMID:25184501; <http://dx.doi.org/10.1038/nbt.3026>
- 38.** Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* 2007; 104:3312-7; PMID:17360644; <http://dx.doi.org/10.1073/pnas.0611511104>
- 39.** Werner T, Koshikawa S, Williams TM, Carroll SB. Generation of a novel wing colour pattern by the Wingless morphogen. *Nature* 2010; 464:1143-8; PMID:20376004; <http://dx.doi.org/10.1038/nature08896>
- 40.** Gloor G, Engels W. Single-fly DNA preps for PCR. *Drosoph Inf Serv* 1992; 71:148-9

Chapter 3

***ebony* Affects Pigmentation Divergence and Cuticular Hydrocarbons in *Drosophila americana* and *D. novamexicana*ⁱ**

Abstract

Drosophila pigmentation has been a fruitful model system for understanding the genetic and developmental mechanisms underlying phenotypic evolution. For example, prior work has shown that divergence of the *tan* gene contributes to pigmentation differences between two members of the virilis group: *Drosophila novamexicana*, which has a light yellow body color, and *D. americana*, which has a dark brown body color. Quantitative trait locus (QTL) mapping and expression analysis has suggested that divergence of the *ebony* gene might also contribute to pigmentation differences between these two species. Here, we directly test this hypothesis by using CRISPR/Cas9 genome editing to generate *ebony* null mutants in *D. americana* and *D. novamexicana* and then using reciprocal hemizyosity testing to compare the effects of each species' *ebony* allele on pigmentation. We find that divergence of *ebony* does indeed contribute to the pigmentation divergence between species, with effects on both the overall body color as well as a difference in pigmentation along the dorsal abdominal midline. Motivated by recent work in *D. melanogaster*, we also used the *ebony* null mutants to test for effects of *ebony* on cuticular hydrocarbon (CHC) profiles. We found that *ebony* affects CHC abundance in both species, but does not contribute to qualitative differences in the CHC profiles between these two species. Additional transgenic resources for working with *D. americana* and *D. novamexicana*, such as *white* mutants of both species and *yellow* mutants in *D. novamexicana*, were

ⁱ This chapter is published as: Lamb, Abigail M., Zinan Wang, Patricia Simmer, Henry Chung, and Patricia J. Wittkopp. 2020. "Ebony Affects Pigmentation Divergence and Cuticular Hydrocarbons in *Drosophila Americana* and *D. Novamexicana*." *Frontiers in Ecology and Evolution* 8: 184.

generated in the course of this work and are also described. Taken together, this study advances our understanding of loci contributing to phenotypic divergence and illustrates how the latest genome editing tools can be used for functional testing in non-model species.

Introduction

Insect pigmentation is a well-studied trait that displays a variety of phenotypic differences within and between species (Wittkopp et al., 2003a; Kronforst et al., 2012). These differences have evolved over a wide range of divergence times and in a great diversity of ecological contexts. Differences in insect pigmentation often appear to be ecologically relevant, correlating with geographic and climatic factors and playing a role in phenomena such as mate recognition, camouflage, thermoregulation, and water balance (True, 2003; Wittkopp and Beldade, 2009). Studies of pigmentation differences within the genus *Drosophila* have emerged as a productive model for studying the evolution of development, exploiting the diversity of phenotypes as well as genetic tools available for working with *Drosophila* and a long history of research into the genetic and biochemical mechanisms controlling pigmentation development (Wittkopp et al., 2003a; Massey and Wittkopp, 2016; Rebeiz and Williams, 2017). Indeed, since the early 2000s, the genetic bases of dozens of pigmentation differences have been identified in varying levels of detail. Strikingly, in every case where a causal role has been directly attributed to a specific gene, the mechanism of change has been found to be a *cis*-regulatory change that affects gene expression rather than a change in the protein's function (Massey and Wittkopp, 2016). These case studies have also identified multiple independent instances of divergent expression for some pigmentation genes, suggesting that these genes are particularly tractable routes for the evolution of pigmentation in this genus (Massey and Wittkopp, 2016).

Changes in *cis*-regulatory sequences are thought to be a common mechanism of developmental evolution because they tend to be less pleiotropic than changes in protein function (Wray et al., 2003; Carroll, 2005). For example, a *cis*-regulatory change

might alter a gene's expression in only a single tissue or a single point in development whereas changing its protein function is expected to impact the organism everywhere that protein is expressed. Genes controlling pigmentation development in *Drosophila* might be especially likely to evolve using this mechanism because the proteins encoded by these genes are also required for other biological functions. For example, genes required for pigment synthesis have also been shown to affect mating success, circadian rhythm, vision, and innate immunity (Nappi and Christensen, 2005; True et al., 2005; Suh and Jackson, 2007; Wittkopp and Beldade, 2009; Takahashi, 2013; Massey et al., 2019a). The pigmentation biosynthesis genes *ebony* and *tan* have also been found to affect the profiles of cuticular hydrocarbons on adult flies, which are hydrophobic lipids on the surface of insect cuticle that are involved in chemical communication, mate recognition, and water balance (Chung et al., 2014; Chung and Carroll, 2015; Massey et al., 2019b).

Here, we investigate genetic changes contributing to the evolution of novel body color in *D. novamexicana*. This species has evolved a much lighter and more yellow body color than its sister species *D. americana* during the approximately 400,000 years since these species diverged from their most recent common ancestor (Figure 3-1; Caletka and McAllister, 2004; Morales-Hojas et al., 2008). *D. novamexicana* and *D. americana* show signs of reproductive isolation (Patterson and Stone, 1949; Ahmed-Braimah and McAllister, 2012), but they are interfertile and can produce viable, fertile F₁ hybrids in the laboratory, allowing genetic analysis (Wittkopp et al., 2003b, 2009). Prior genetic mapping has identified two quantitative trait loci (QTL) that together account for ~87% of the pigmentation difference between *D. novamexicana* and *D. americana* (Wittkopp et al., 2009). Fine mapping and transgenic analysis revealed that the QTL of smaller effect was driven by divergence at *tan* (Wittkopp et al., 2009), a gene that encodes a hydrolase that catalyzes the conversion of N-B-alanyl dopamine (NBAD) to dopamine, a precursor for dark melanin pigment (True et al., 2005). The QTL of larger effect was linked to an inverted region containing the candidate gene *ebony*, but the presence of the inversion prevented fine mapping to separate the effects of *ebony* from linked loci (Wittkopp et al., 2009). *ebony* encodes a synthetase that catalyzes the conversion of dopamine into NBAD, a precursor for light yellow pigments

(Koch et al., 2000), which is the opposite of the reaction catalyzed by Tan. *ebony* has also been shown to have expression differences between *D. novamexicana* and *D.*

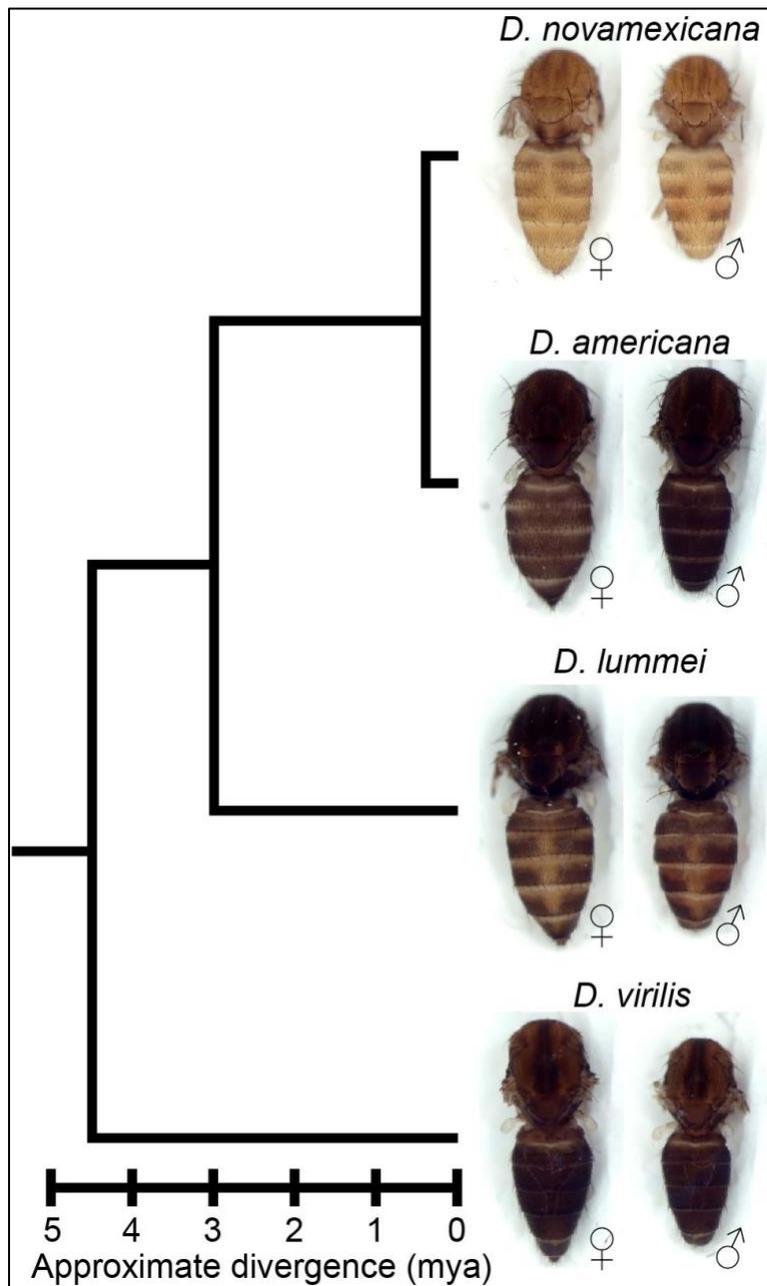


Figure 3-1: *Drosophila novamexicana* shows divergent body color within the virilis group. Phylogenetic relationships with estimated divergence times (Caletka and McAllister, 2004; Cooley et al., 2012) are shown for *D. novamexicana*, *D. americana*, *D. lummei*, and *D. virilis*. For each species, a dorsal view of the thorax and abdomen is shown for females (left) and males (right), with heads, wings, and legs removed.

americana caused by *cis*-regulatory divergence (Cooley et al., 2012).

Despite these data suggesting that *ebony* contributes to pigmentation divergence between *D. novamexicana* and *D. americana*, the phenotypic effects of sequence divergence at *ebony* have not been demonstrated. Here, we show that divergence at *ebony* does indeed contribute to pigmentation divergence between these two species. We use CRISPR/Cas9 genome editing to mutate *ebony* in both species and use these mutant genotypes to directly test *ebony*'s contribution to pigmentation divergence through reciprocal hemizyosity testing (Stern, 2014). We find that the *D. novamexicana* *ebony* allele causes lighter pigmentation throughout the

body than the *D. americana ebony* allele. We also find that allelic divergence at *ebony* is primarily responsible for a spatial difference in abdominal pigmentation between these species: the *D. novamexicana ebony* allele causes the absence of dark melanin along the dorsal midline of the abdomen seen in *D. novamexicana*. Finally, we show that *ebony* affects the cuticular hydrocarbon (CHC) profiles in *D. americana* and *D. novamexicana*, but does not contribute to the qualitative differences in CHC profiles seen between species. Taken together, our data show the power of using CRISPR/Cas9 genome editing to test functional hypotheses about evolutionary mechanisms. In addition, resources generated and lessons learned in the course of this work are expected to help other researchers perform CRISPR/Cas9 genome editing in *D. americana*, *D. novamexicana* and other *Drosophila* species.

Materials and Methods

Fly stocks and husbandry

The following fly lines were used in this study: *D. americana* “A00” (National Drosophila Species Stock Center number 15010-0951.00), *D. novamexicana* “N14” (National Drosophila Species Stock Center number 15010-1031.14), *D. lummei* (National Drosophila Species Stock Center number 15010-1011.08), *D. virilis* (National Drosophila Species Stock Center number 15010-1051.87), *D. melanogaster* $y^1 M\{w[+mC]=nos-Cas9.P\}ZH-2A w^*$ (Bloomington Drosophila Stock Center number 54591), and *D. melanogaster* Canton-S. All flies were reared on standard cornmeal medium at 23–25°C with a 12:12 hour light:dark cycle.

Transgenesis and CRISPR mutant generation in D. americana and D. novamexicana

To the best of our knowledge, prior to this work, the only transformation of *D. americana* or *D. novamexicana* resulted from the insertion of a piggyBac transgene (Wittkopp et al., 2009). We therefore first used the CRISPR/Cas9 system to generate *white* mutants in both species to test the feasibility of CRISPR genome modification and to create lines that are easier to screen for common transformation

markers that drive expression of fluorescent proteins or restore red pigmentation in the eyes by restoring *white* function. We successfully generated *white* mutant N14 and A00 lines used as transgenic hosts for future work, by injecting single guide RNAs (sgRNAs) targeting coding sequences in *white* conserved between *D. novamexicana* and *D. americana* in the second and third exons and screening for the loss of red eye pigment in male offspring of injected females (Supplemental Figure S3-1); *white* is on the X chromosome and thus only present in a single copy in males. These same guide RNAs were also used in *Drosophila virilis* to cut *white* and integrate an attP landing site potentially useful for site-directed transgene insertion (Lachowiec and Wittkopp, unpublished data), although the PhiC31 system does not seem to work well in *D. virilis* (Stern et al., 2017). For all CRISPR experiments, sgRNAs were *in vitro* transcribed from DNA templates using Invitrogen T7 MEGAscript Transcription Kit according to protocol described by Bassett et al. (2013). Oligonucleotides used to generate sgRNAs are listed in Supplemental Table S3-1. After transcription, sgRNAs were purified using RNA Clean and Concentrator 5 kit (Zymo Research), eluted with nuclease-free water, and quantified with Qubit RNA BR Assay Kit (Thermo Fisher Scientific). For CRISPR injections, sgRNAs were mixed with purified Cas9 protein (PNA Bio #CP01) with a final injected concentration of 0.05% phenol red to visualize the injection mix. CRISPR injections were performed in-house, using previously described methods (Miller et al., 2002).

To try to increase efficiency of CRISPR mutagenesis in these species, we next sought to generate transgenic lines expressing Cas9 in the germlines of *white* mutant *D. americana* (A00) and *D. novamexicana* (N14) flies using piggyBac transgenesis (Horn and Wimmer, 2000). Based on prior reports that the *nanos* (*nos*) promoter and 3'UTR drive expression in the germline of *Drosophila virilis* (Holtzman et al., 2010), a close relative of *D. americana* and *D. novamexicana*, we amplified the *nos-Cas9-nos* transgene from the pnos-Cas9-nos plasmid (Addgene #62208; Port et al., 2014) using Phusion High Fidelity Polymerase (NEB) with tailed primers and cloned the amplicon into pBac{3XP3-ECFPafm} (Horn and Wimmer, 2000) digested with *Ascl* and *Bsu36l* restriction enzymes using Gibson Assembly Master Mix (NEB). Primers are included in Supplemental Table S3-1. We confirmed the insert was correctly

incorporated and free of PCR-induced errors by Sanger sequencing. We sent the *white* mutant lines of *D. americana* A00 and *D. novamexicana* N14 that we generated to Rainbow transgenic services for piggyBac transgenesis (www.rainbowgene.com) and screened offspring of injected adults for expression of the enhanced cyan fluorescent protein (ECFP) in the eye using a Leica MZ6 stereoscope equipped with a Kramer Scientific Quad Fluorescence Illuminator. Transformants were obtained from injections into *D. novamexicana* (N14) (PCR verified), but not from injections into *D. americana* (A00), despite multiple attempts.

All subsequent CRISPR injections in *D. novamexicana* were performed using flies homozygous for the *nos-Cas9-nos* transgene, some with and some without the inclusion of commercially available Cas9 protein in the injection mix. CRISPR mutants were only obtained from injections containing the commercially available Cas9 protein, however, suggesting that the *nos-Cas9-nos* transgene might not drive expression of Cas9 in the germline of *D. novamexicana*. To test this hypothesis, we used western blotting to examine Cas9 protein expression in 3 transformed *D. novamexicana* N14 lines with independent insertions of the piggyBac transgene and in a *D. melanogaster* transgenic line carrying the original *pnos-Cas9-nos* transgene (Bloomington Drosophila Stock Center line 54591, transformed with Addgene plasmid #62208, Port et al., 2014). These experiments showed that the *nos-Cas9-nos* transgene in *D. novamexicana* N14 flies does not express Cas9 protein in the ovaries (Supplemental Figure S3-2). This conclusion was further supported when injection of sgRNAs targeting the *yellow* gene into the *D. novamexicana* line carrying the *nos-Cas9-nos* transgene also only produced *yellow* mutants when the Cas9 protein was co-injected with the sgRNAs (Supplemental Figure S3-3). Ability of the *nos* promoter to drive germline expression in the closely related species *D. virilis* has also been found to be variable among transgenic lines (Hannah McConnell, Aida de la Cruz, and Harmit Malik, personal communication), suggesting that other promoters should be used in the future to drive reliable germline expression in the virilis group.

To generate *ebony* mutant *D. americana* (A00) and *D. novamexicana* (N14), we synthesized five sgRNAs targeting conserved sites in the first coding exon of *ebony*.

Because *ebony* is located on an autosome and *ebony* loss-of-function mutant alleles are generally considered recessive in *D. melanogaster* (Thurmond et al., 2019), we did not expect to be able to identify *ebony* mutants by simply screening progeny of injected flies for mutant phenotypes as we did for *white* and *yellow*. We therefore co-injected a donor plasmid containing the sequence of an eye-specific red fluorescent protein marker (3XP3-RFP) flanked by *ebony* sequences that could be inserted into *ebony* via homology-directed repair and used to screen for *ebony* mutants. Although we observed RFP expression in larvae injected with the homology-directed repair donor fragment, indicating that the reporter gene was functional in these species, injected individuals did not produce any offspring with red fluorescent eyes, suggesting that the donor plasmid was not integrated in the germline of injected individuals. Because non-homologous end joining occurs more frequently than homology directed repair following double-strand breaks (Liu et al., 2018), we also tried to identify flies that might be heterozygous for an *ebony* mutant allele by closely inspecting all offspring of injected (G_0) flies for any subtle changes in pigmentation. Specifically, we collected and mated (G_1) offspring of injected flies with any noticeably darker pigmentation, keeping them grouped by G_0 parent of origin. As further described in the results, we were ultimately able to identify homozygous *ebony* mutants among progeny from these $G_1 \times G_1$ crosses of relatively dark flies derived from two independent *D. novamexicana* G_0 flies and one *D. americana* G_0 fly. Sanger sequencing these flies confirmed they were homozygous for *ebony* alleles containing deletions. We then crossed the mutated *ebony* alleles back into wild-type backgrounds of each parental species to generate homozygous *ebony* mutant lines with wild-type red eyes.

Western blotting

For *ebony* western blotting, proteins were extracted from stage P14/15 pupae, identified by the following characteristics: black pigmentation present in wings and bristles, meconium visible in abdomen (Cooley et al., 2012). For each sample, five pupae were homogenized in 100 μ L of homogenization buffer (125 mM Tris pH 6.8, 6% SDS, 2.5X Roche cOmplete protease inhibitor cocktail, EDTA-free), then centrifuged for

15 min at 15000 rcf, and the supernatant transferred to a fresh tube with an equal volume of 2x Laemmli buffer (125 mM Tris pH 6.8, 6% SDS, 0.2% glycerol, 0.25% bromophenol blue, 5% Beta-mercaptoethanol).

For Cas9 western blotting, protein was extracted from ovaries dissected in ice cold PBS from the following lines: untransformed N14 *white* mutants (host line), three independently transformed lines of N14 *white* carrying the pBac{3XP3-ECFPafm-nosCas9nos} transgene, transgenic *D. melanogaster* carrying the *pnos-Cas9-nos* transgene, and wild-type (Canton-S) *D. melanogaster*. For *D. novamexicana* samples, we collected ovaries from 10 sexually mature flies, whereas for *D. melanogaster* samples, we collected ovaries from 18 sexually mature flies. Different numbers of flies were used for the two species because of differences in body size. In each case, ovaries were placed into microcentrifuge tubes on ice, spun down briefly in a tabletop centrifuge, and excess PBS was removed and replaced with 20 uL of homogenization buffer. Samples were then treated as described for *ebony* western blots above. A positive control Cas9 sample was made by diluting purified Cas9 protein (PNA Bio CP01) in homogenization buffer, and mixing with 2X Laemmli buffer to a final concentration of 2.5 ng/uL.

Samples were heated at 95°C for 10 min before loading into 7.5% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) and running at 150V for approximately 90 min at 4°C in 1X tris-glycine running buffer. Separate gels were run for *ebony* and Cas9 blots. Samples were loaded in the following volumes: 35 uL per pupa sample, 30 uL per ovary sample, 10 uL of Cas9 positive control (25 ng protein), 5 uL PageRuler prestained protein ladder (Thermo Fisher Scientific). Gels were transferred onto PVDF membrane in tris-glycine transfer buffer, 10% MeOH, 0.01% SDS at 100 V for 1 h with stirring on ice at 4°C. Membranes were blocked in 3% nonfat dry milk in TBST for 30 min at RT with shaking, then divided in half using the prestained ladder as a guide just below the 100 kDa mark for the Cas9 membrane and just below the 70 kDa mark for the *ebony* membrane. The lower molecular weight halves of the membranes were placed in solutions containing primary antibodies to detect the protein used as a loading control (tubulin or lamin), whereas the halves of the membranes

containing the higher molecular weight proteins were placed in solutions containing primary antibody solutions against the protein of interest (Ebony or Cas9), each diluted in 3% nonfat dry milk in TBST. In all cases, membranes were incubated with the primary antibodies overnight at 4°C. Primary antibody solutions for *ebony* included rabbit anti-*ebony* 1:300 (Wittkopp et al., 2002) and rabbit anti-alpha tubulin 1:5000 (Abcam ab52866) as a loading control. Primary antibody solutions for Cas9 included mouse anti-Cas9 1:1000 (Novus NBP2-36440) and mouse anti-lamin 1:200 (DHSB adl67.10) as a loading control. Membranes were washed in TBST and transferred to secondary antibody solutions diluted in 3% nonfat milk in TBST for 2 h at RT. The following secondary antibodies were used: donkey anti-rabbit HRP 1:5000 (Amersham na934) or goat anti-mouse HRP 1:5000 (abcam ab97023). Membranes were finally washed in TBST and developed with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) and imaged using a Licor Odyssey FC imaging system.

Fly crosses for reciprocal hemizyosity testing and cuticular hydrocarbon analysis

To generate F₁ hybrids carrying only one (*D. americana* or *D. novamexicana*) functional *ebony* allele, wild-type and *ebony* mutant flies from each species were collected as virgins and aged in vials for at least 12 days to reach sexual maturity and verify virgin female status by absence of larvae. Crosses were all set on the same batch of food on the same day and placed at 25°C. For most crosses, 4 virgin females and 4 males were used; however, 8 virgin females and 8 males were used in interspecific crosses with *D. novamexicana* females because of reduced mating success in these crosses. After 3 days, adult flies from these crosses that would be used for cuticular hydrocarbon (CHC) analysis were transferred to new vials with a fresh batch of food. Offspring from the first set of vials were used for imaging and pigmentation analysis, while offspring from the second set of vials were used for CHC analysis. Flies used for pigmentation phenotyping were aged 5–7 days after eclosion and preserved in 10% glycerol in ethanol before imaging (Wittkopp et al., 2011).

Imaging of fly phenotypes

Insect specimens were imaged using a Leica DC480 camera attached to a Leica MZ16F stereoscope equipped with a ring light attachment and Leica KL 1500 LCD lamp. Images were captured using Leica DC Twain software version 5.1.1 run through Adobe Photoshop CS6 version 13.0 X32. Prior to imaging, pupal cases and wings were mounted on slides in PVA mounting medium (BioQuip). Thorax, abdomen, and whole-body specimens were prepared from age-matched, preserved flies as described in the previous section. For imaging, thorax, abdomen, and whole-body specimens were submerged in 100% ethanol in custom wells composed of white oven-cured polymer clay (Sculpey).

Because the color of specimens spanned a wide range across genotypes, exposure was optimized for each sample type (e.g., whole body, thorax, abdomen, wing, pupal case) individually by placing specimens from the two phenotypic extremes in the same frame and adjusting exposure to avoid over-exposing the lightest flies while capturing as much detail as possible from the darkest flies. Exposure time, lighting, white balance, background, and zoom were kept identical across all images of single tissue type. Minor color adjustments to improve visibility of phenotypes were performed simultaneously across all raw images of the same sample type in a single combined document using Photoshop CC 2019, ensuring that all images presented for direct comparisons were adjusted identically.

Cuticular hydrocarbon analyses

CHCs for each cross were extracted from 5-day-old females by soaking the flies for 10 min in 200 μ l hexane containing hexacosane (C26; 25 ng/ μ l) as an internal standard. Eight replicates were prepared for each cross. Extracts were directly analyzed by the GC/MS (7890A, Agilent Technologies Inc., Santa Clara, CA, United States) coupled with a DB-17ht column 30 m by 0.25 mm (i.d.) with a 0.15 μ m film thickness (Agilent Technologies Inc., Santa Clara, CA, United States). Mass spectra were acquired in Electron Ionization (EI) mode (70 eV) with Total Ion Mode (TIM) using the GC/MS (5975C, Agilent Technologies Inc., Santa Clara, CA, United States). The peak areas were recorded by MassHunter software (Agilent Technologies Inc., Santa Clara, CA, United States). Helium was the carrier gas at 0.7 ml/min and the GC thermal

program was set as follows: 100°C for 4 min, 3°C/min to 325°C. Straight-chain compounds were identified by comparing retention times and mass spectra with authentic standard mixture (C6-C40) (Supelco® 49452-U, Sigma-Aldrich, St. Louis, MO, United States). Methyl-branched alkanes, alkenes, dienes and trienes were then identified by a combination of their specific fragment ions on the side of functional groups (methyl branch or double bonds) and retention times relative to linear-chain hydrocarbon standards. Each individual CHC peak was quantified by normalizing its peak area to the peak area of the internal C26 standard, converting each CHC peak area to ng/fly using the known internal standard concentration of 1000 ng/fly. Welch's *t*-tests with a Benjamini–Hochberg correction for multiple testing (Benjamini and Hochberg, 1995) were used to compare CHC amounts between pairs of genotypes. Because the effect of *ebony* on individual CHC abundance in *D. melanogaster* was recently shown to increase with CHC chain length (Massey et al., 2019b), we also compared the effects of *ebony* loss of function on different chain-lengths of CHCs. Eight biological replicates of homozygous *ebony* null measurements were divided by the mean measurement of the eight replicates of the matched *ebony* heterozygote for each individual CHC. The ratio of *ebony* null to heterozygote CHC abundance was plotted against CHC chain length. The relative effects of *D. americana* versus *D. novamexicana* *ebony* in a common F₁ hybrid background (described as F₁[*e^A/e⁻*] and F₁[*e^N/e⁻*], respectively), were also compared in this manner, with the replicates of the F₁[*e^A/e⁻*] divided by the mean F₁[*e^N/e⁻*] measurement for each CHC. We used Spearman's rank correlation (Spearman's rho) to test the relationship between CHC chain length and the effect of *ebony* on CHC abundance. The threshold for statistical significance was set at *alpha* = 0.05 for all tests. Datafile and R code used for this analysis are provided in Supplemental File S3-1 and Supplemental File S3-2, respectively.

Results and Discussion

The reciprocal hemizyosity test is a powerful strategy for identifying genes with functional differences that contribute to phenotypic divergence [reviewed in Stern

(2014)]. This test is performed by comparing the phenotypes of two hybrid genotypes that are genetically identical except for which allele of the candidate gene is mutated. Any phenotypic differences observed between these two genotypes are attributed to divergence of the candidate gene. Applying this test to identify functional differences between species requires loss-of-function (null) mutant alleles in both species and the ability for the species to cross and produce F₁ hybrids. Consequently, in order to use this strategy to test *ebony* for functional divergence between *D. novamexicana* and *D. americana*, we first needed to generate *ebony* null mutant alleles in both species.

Generating *ebony* mutants in *D. americana* and *D. novamexicana* using CRISPR/Cas9

We generated *ebony* null mutants in *D. novamexicana* and *D. americana* by using CRISPR/Cas9 to target double-strand breaks to five conserved sites within the first coding exon of *ebony*. As described more fully in the “Materials and Methods” section, we injected embryos of *white* mutants from both species with purified Cas9 protein and sgRNAs targeting all five sites simultaneously. BLAST searches showed that all of the sgRNAs targets were at least 5 bp different from all other sequences in genomes from two different strains of *D. americana*. Prior work in *D. melanogaster* has shown that heritable off-target mutations were never recovered in sequences with 3 or more mismatches to the sgRNA (Ren et al., 2014). To make it easier to identify *ebony* mutant alleles, we also injected a donor plasmid that would allow homology directed repair to integrate a transgene expressing red fluorescent protein in the fly’s eyes, but no progeny of injected flies were observed to express this transformation marker. However, we reasoned that although we were unable to insert a marker at *ebony*, the CRISPR machinery may still have induced double-strand breaks in the target sequence, and *ebony* mutants could have been generated by non-homologous end-joining resulting in deletions or insertions. Therefore, we also searched for *ebony* mutants by looking for changes in body pigmentation.

In *D. melanogaster*, *ebony* loss-of-function mutants have a much darker appearance than wild-type flies because they are unable to produce yellow sclerotin, causing an increase in production of black and brown melanins (Wittkopp et al., 2002). *D.*

melanogaster ebony mutant alleles are commonly described as recessive to wild-type *ebony* alleles (Thurmond et al., 2019); however, in some genetic backgrounds, flies heterozygous for an *ebony* mutant allele are slightly darker than wild-type flies (Thurmond et al., 2019). Because *D. novamexicana* has such a light yellow body color (Figure 3-2A), we thought it possible that flies heterozygous for an *ebony* mutant allele might also show a detectable darkening of pigmentation; we were less optimistic about being able to detect heterozygous *ebony* mutants based on pigmentation in *D. americana* because its wild-type pigmentation is already very dark (Figure 3-2C). Nonetheless, we sorted through the progeny of

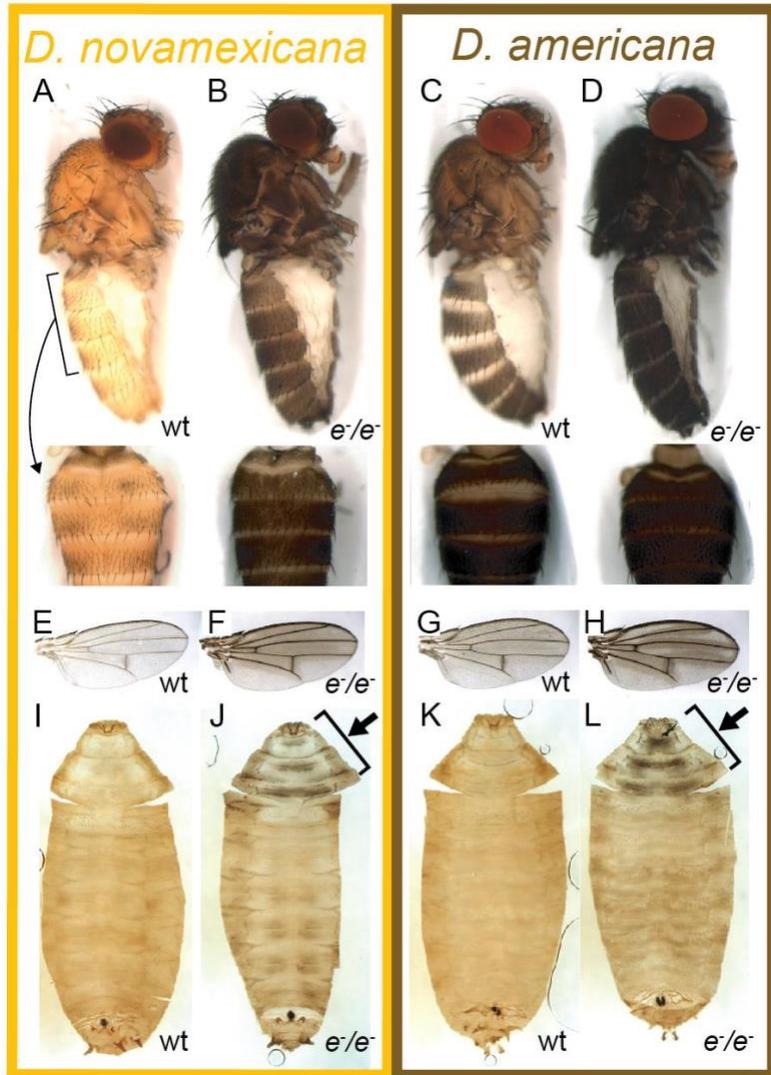


Figure 3-2: Ebony affects body, wing, and pupal pigmentation in *D. novamexicana* and *D. americana*. (A–D) Adult body pigmentation is shown from a lateral view (top) and dorsal abdominal view (segments A2–A4, bottom) for (A) *D. novamexicana*, (B) *D. novamexicana ebony* null mutants, (C) *D. americana*, and (D) *D. americana ebony* null mutants. (E–H) Adult wing pigmentation is shown for (E) *D. novamexicana*, (F) *D. novamexicana ebony* null mutants, (G) *D. americana*, and (H) *D. americana ebony* null mutants. (I–L) Pigmentation of pupal cases is shown for (I) *D. novamexicana*, (J) *D. novamexicana ebony* null mutants, (K) *D. americana*, and (L) *D. americana ebony* null mutants. Arrows in (J,L) highlight the most prominent areas with dark pigmentation in *ebony* mutants.

injected *D. novamexicana* and *D. americana* flies, isolating any individuals that seemed to have darker pigmentation than their siblings and allowing these relatively dark flies to freely mate in vials segregated by injected parents, keeping individual “founder” mutations separate.

Two of the vials of darker pigmented *D. novamexicana* flies produced pupae with an unusual black pattern on the anterior end of the pupal case (Figure 3-2J). We moved these pupae to new vials and found that black-patterned pupae from both “founder” vials developed into adults with the much darker than wild-type body color expected for homozygous *ebony* mutants in *D. novamexicana* (Figures 3-2A,B). Because pigmentation of the pupal case is very similar between *D. novamexicana* and *D. americana* (Figures 3-2I,K, Ahmed-Braimah and Sweigart, 2015), we also searched for pupae with similar pigmentation marks in the vials containing progeny of darker flies descended from injected *D. americana*. We found such pupae in only one of the *D. americana* vials (Figure 3-2L). Flies emerging from these pupal cases also showed darker pigmentation than wild-type *D. americana* (Figures 3-2C,D), as expected for homozygous *ebony* mutants, but this difference was much more subtle than in *D. novamexicana* (Figures 3-2A,B). Flies from both species emerging from pupal cases with abnormal pigmentation also showed increased levels of dark melanins in wings in a pattern similar to that seen in *D. melanogaster ebony* mutants (Figures 3-2E–H, Wittkopp et al., 2002), further suggesting that they were homozygous for *ebony* mutant alleles. Crossing putative homozygous *ebony* mutants from the same species to each other resulted in true-breeding lines of *D. americana* and *D. novamexicana* presumed to be homozygous for *ebony* mutant alleles.

To determine whether these true-breeding lines were indeed homozygous for *ebony* mutant alleles, we used Sanger sequencing to search for changes in the *ebony* sequence in the region targeted for double strand breaks with CRISPR/Cas9. We found that the presumed *ebony* mutant lines of both species harbored deletions corresponding to the locations of sgRNA target sites in the first coding exon, with the two *D. novamexicana* mutant lines carrying deletions of 7 and 10 bases and the *D. americana* mutant line carrying a deletion of 46 bases (Figure 3-3A). Each of these

mutations is expected to cause frameshifts, leading to multiple early stop codons. Further experiments described in this study using *D. novamexicana ebony* mutants were conducted with the 10 base deletion line, and any further description of *ebony* null *D. novamexicana* refers to this line.

To further assess whether these mutations caused null alleles, we used western blotting to examine the expression of the Ebony protein during late pupal stages when adult pigmentation is developing and the *ebony* gene is expressed in the developing abdomen (Wittkopp et al., 2002; Cooley et al., 2012). We performed western blots on protein extracts from P14/P15 stage pupae of both wild-type and homozygous *ebony* mutant flies of both *D. americana* and *D. novamexicana* using an antibody against *D. melanogaster ebony* (Wittkopp et al., 2002). This antibody recognizes a 94 kDa protein consistent with the predicted molecular weight of Ebony in pupal protein extracts from wild-type lines of both *Drosophila melanogaster* and *Drosophila biarmipes*, but does not produce a 94 kDa band in pupal protein extracts of either *e1* or *In(3R)eAFA ebony* mutant lines of *D. melanogaster* (Wittkopp et al., 2002). Wild-type extracts of both *D. americana* and *D. novamexicana* produced presumptive Ebony bands while extracts from flies homozygous for *ebony* deletions did not produce a 94 kDa band for either species (Figure 3-3B). The nature of the frameshift deletions as well as the western blot evidence together show that these *ebony* mutations cause null alleles.

ebony* divergence contributes to body color differences between *D. novamexicana* and *D. americana

We used the homozygous *ebony* mutant *D. novamexicana* and *D. americana* lines to perform a reciprocal hemizygoty test by crossing *ebony* mutant *D. novamexicana* (e^-/e^-) to wild-type *D. americana* (e^A/e^A) and *ebony* mutant *D. americana* (e^-/e^-) to wild-type *D. novamexicana* (e^N/e^N) (Figure 3-4A). In order to observe the effects of the two species' *ebony* alleles in the presence of each species X chromosome, we conducted sets of reciprocal crosses (i.e., swapping the genotypes of the male and female parents). Female F₁ hybrids from reciprocal crosses are genetically

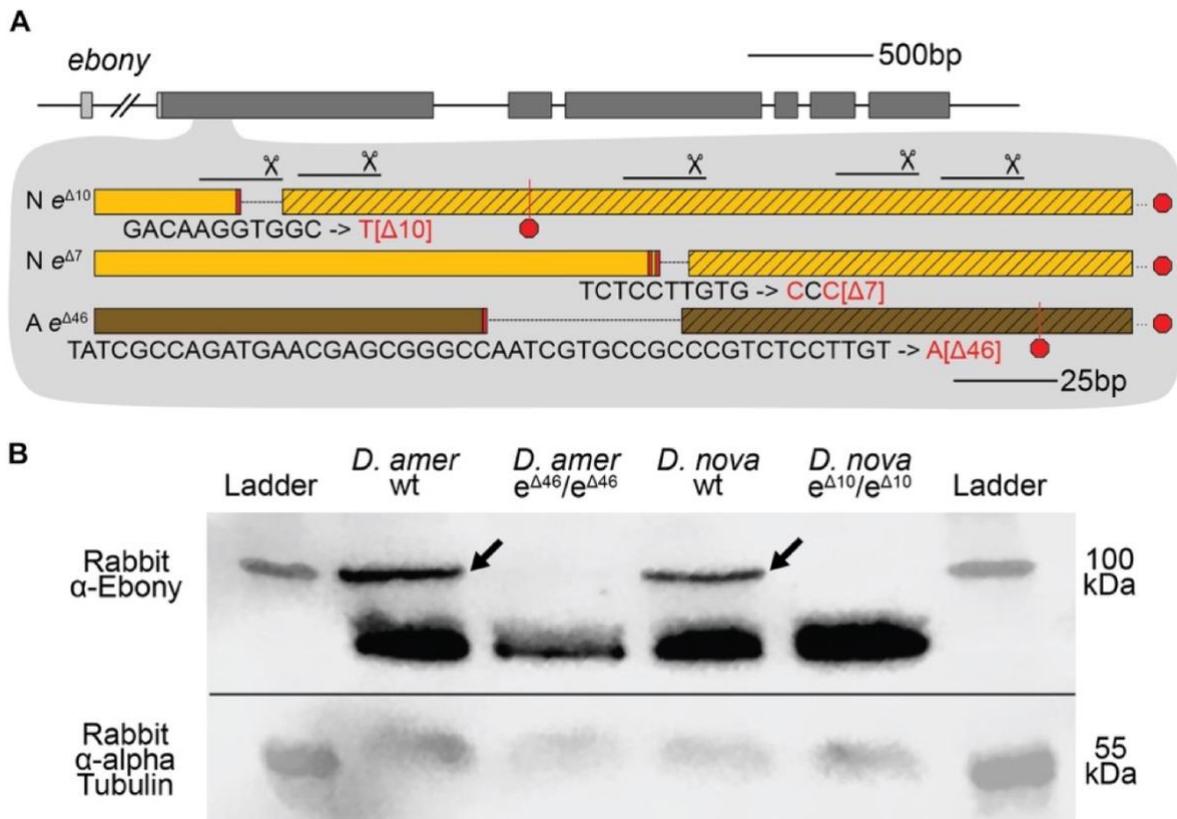


Figure 3-3: CRISPR/Cas9-induced mutations created null alleles of the *D. novamexicana* and *D. americana* *ebony* genes. (A) A schematic of the *ebony* gene is shown with gray boxes indicating exons; coding sequence is indicated in the darker shade of gray. Locations of the five guide RNAs targeting the second exon of *ebony* are shown with solid lines below scissor symbols. Mutations observed in the two *ebony* mutants ($e^{\Delta 10}$ and $e^{\Delta 7}$) isolated in *D. novamexicana* (“N”) and the one *ebony* mutant ($e^{\Delta 46}$) isolated in *D. americana* (“A”) are shown. All three alleles included deletions that caused frameshifts. (B) Western blotting showed that the *D. americana* $e^{\Delta 46}$ and *D. novamexicana* $e^{\Delta 10}$ mutants (lanes 2 and 4, respectively), lacked a ~100 kDa protein (arrows) recognized by an antibody raised against *D. melanogaster* Ebony protein (Wittkopp et al., 2002) that is present in wild-type (wt) *D. americana* and *D. novamexicana* (lanes 1 and 3, respectively). Relative abundance of total protein loaded into each lane can be seen by the relative intensities of the shorter proteins also detected by the Ebony antibody (Wittkopp et al., 2002) as well as the relative intensities of ~55 kDa bands detected by an antibody recognizing alpha Tubulin (Abcam ab52866). The solid black line shows where the membrane was cut prior to incubation with primary antibodies during the western blotting procedure; the top half was incubated with anti-Ebony antibodies whereas the bottom half was incubated with anti-Tubulin antibodies. The two halves were realigned by hand for imaging, using the shape of the cut and the ladder staining as a guide. An unannotated image of this blot is shown in Supplemental Figure S3-4.

identical except for the parent of origin of their one functional *ebony* allele (e^N or e^A). F₁ hybrid females carrying a functional *D. novamexicana* *ebony* allele (F₁[e^N/e^-]) developed a lighter body color than F₁ hybrid females carrying a functional *D. americana* *ebony* allele (F₁[e^A/e^-]) (Figures 3-4B,C vs. Figures 3-4D,E). These data demonstrate for the first time that functional divergence between the *D.*

novamexicana and *D. americana* *ebony* alleles contributes to divergent body color between these two species.

To determine how *ebony* divergence interacts with divergent loci on the X-chromosome, we also compared the body color of male progeny from these reciprocal crosses. Like the F₁ hybrid females, these F₁ hybrid males differ for the parent of origin for their one functional *ebony* allele (e^A or e^N); however, they also differ for the parent of origin of all X-linked genes. Prior work has shown that divergence on the X-chromosome, particularly divergence in non-coding sequences of the *tan* gene, also contributes to differences in body color between *D. novamexicana* and *D. americana* (Wittkopp et al., 2003b, 2009). As expected, we found that body color differed between males carrying alternate species' X chromosomes (Figure 3-4F vs. Figure 3-4G and Figure 3-4H vs. Figure 3-4I) as well as between males carrying the same X chromosome but different species' functional *ebony* alleles (Figure 3-4F vs. Figure 3-4H and Figure 3-4G vs. Figure 3-4I). Consistent with prior findings demonstrating that divergence in the QTL containing *ebony* explained more of the difference in pigmentation than divergence at X-linked genes, we found that males with functional *D. americana* *ebony* alleles had the darkest phenotypes, regardless of their X-chromosome genotype (Figures 3-4F–I).

ebony* divergence also contributes to a difference in abdominal pigment patterning between *D. novamexicana* and *D. americana

Although the divergent overall body color is the most striking difference in pigmentation between *D. novamexicana* and *D. americana*, there is also a difference in the distribution of pigments along the dorsal midline of the abdomen between these two species (Figure 3-1). This difference is also visible in individuals of both species heterozygous for an *ebony* null allele (Figures 3-4J–M). Prior work has shown that the absence of dark pigments seen in this region of *D. novamexicana* is dominant in F₁ hybrids to the presence of dark pigments seen in this region of *D.*

americana (Wittkopp et al., 2003b). In addition, genetic mapping of this trait between *D. novamexicana* and *D. virilis* (which has a dark midline region similar to *D. americana*)

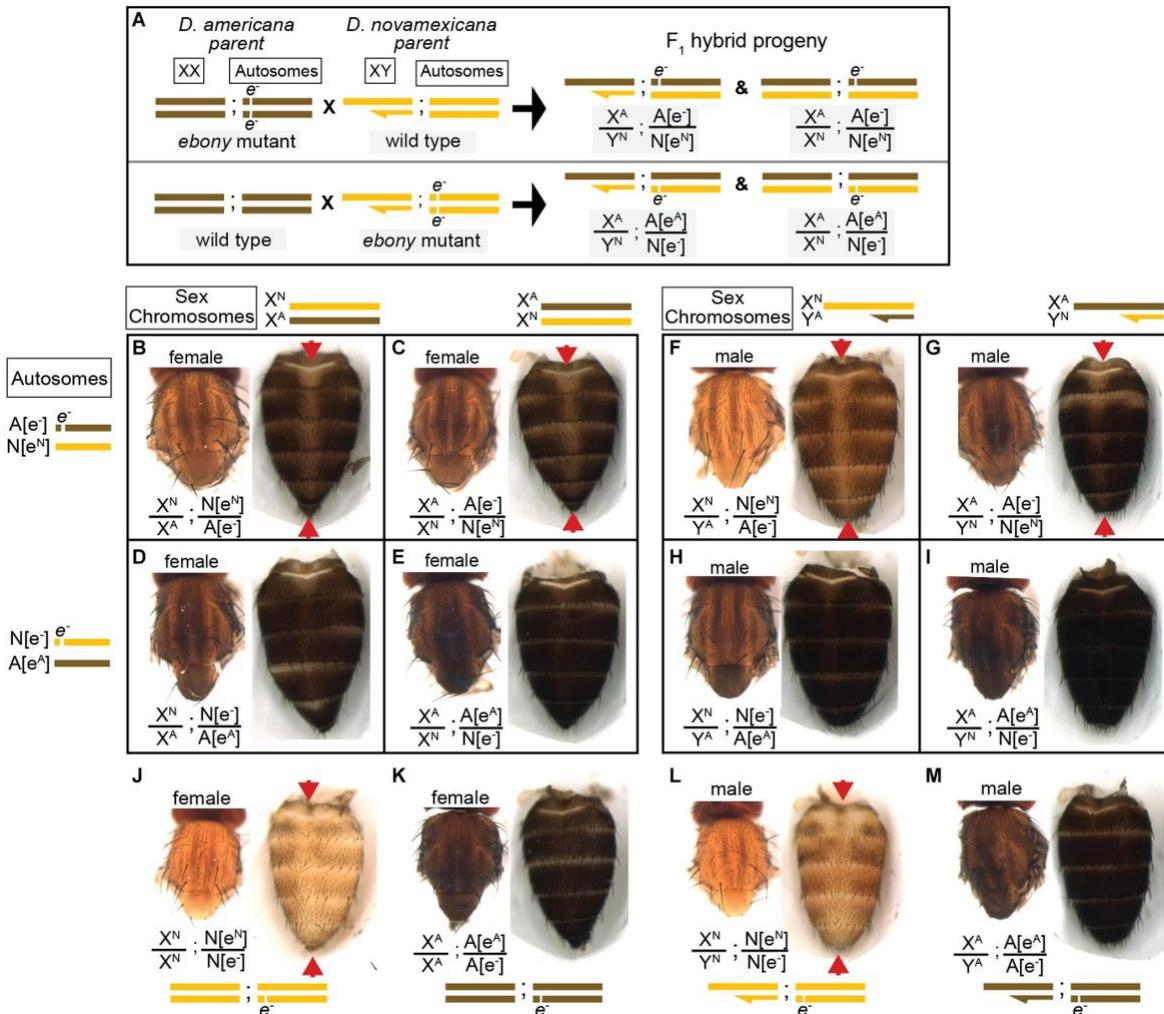


Figure 3-4 Reciprocal hemizyosity testing shows effects of *ebony* divergence between *D. americana* and *D. novamexicana* on body pigmentation. (A) Schematic shows representative sex chromosomes (XX and XY) and autosomes of the parents and progeny of reciprocal hemizyosity crosses, along with the genotypes of the progeny. Although a single autosome is shown for simplicity, these species have five autosomes. Superscript “A” and “N”, as well as brown and yellow colored bars, indicate alleles and chromosomes from *D. americana* and *D. novamexicana*, respectively; e^- indicates an *ebony* null allele. Although the schematic illustrates the crosses only with *D. americana* as the female parent, the same crosses were performed with sexes of the parental species reversed. **(B–I)** Dorsal thorax and abdomen phenotypes are shown for female **(B–E)** and male **(F–I)** progeny of reciprocal hemizyosity crosses. Genotypes of autosomal and sex chromosomes are shown to the left and above panels **(B–I)**, respectively, using the same schematic notation as in panel **(A)**. Individuals in panels **(B,C,F,G)** carry a wild-type copy of *D. novamexicana* *ebony* allele, whereas individuals in panels **(D,E,H,I)** carry a wild-type copy of the *D. americana* *ebony*. **(J–M)** Dorsal thorax and abdomen phenotypes are shown for female **(J,K)** and male **(L,M)** flies heterozygous for the *ebony* null allele in *D. novamexicana* **(J,L)** and *D. americana* **(K,L)** for comparison to flies shown in panels **(B–I)**, which also all carry one null and one wild-type *ebony* allele. Red arrowheads in panels **(B,C,F,G,J,L)** highlight the reduced dark pigmentation in the abdomen along the dorsal midline relative to lateral regions.

has shown that the chromosome including *ebony* (chromosome 2) has a large effect on this trait (Spicer, 1991). We found that *D. novamexicana* *ebony* mutants showed even pigmentation across the width of each abdominal segment (Figure 3-2B), demonstrating that *ebony* is required for the development of lighter pigmentation along the dorsal midline in wild-type *D. novamexicana* (Figure 3-2A). In addition, comparing the pigmentation of this abdominal dorsal midline region between F₁ hybrid flies of both sexes from the reciprocal crosses described above (Figure 3-4) showed that divergence at *ebony* contributes to this trait difference between *D. novamexicana* and *D. americana*. Specifically, we observed less dark pigments in the dorsal midline region of the abdomen in F₁ hybrid individuals inheriting the wild-type *D. novamexicana* *ebony* allele (F₁[*e^N/e⁻*], Figures 3-4B,C,F,G) than the *D. americana* *ebony* allele (F₁[*e^A/e⁻*]) (Figures 3-4D,E,H,I). Males carrying a functional *D. novamexicana* *ebony* allele (F₁[*e^N/e⁻*]) showed reduced pigmentation in the dorsal midline relative to the lateral regions regardless of the origin of their X chromosome (Figures 3-4F,G), indicating that divergent loci on the X-chromosome (including *tan*) do not affect the presence of this phenotype.

Cuticular hydrocarbon profiles differ between *D. americana* and *D. novamexicana* and are affected by *ebony* expression but not *ebony* divergence

ebony expression was recently found to affect the relative abundance of cuticular hydrocarbons (CHCs) in *D. melanogaster* (Massey et al., 2019b). In addition, variation in *ebony* expression was also shown to correlate with variation in CHC profiles among natural isolates of *D. melanogaster* (Massey et al., 2019b). CHC profiles have been shown to vary among virilis group species as well, including between *D. novamexicana* and the strain of *D. americana* used in this study (Bartelt et al., 1986). We therefore asked whether differences in *ebony* might contribute to differences in CHC profiles between these two species using a reciprocal hemizyosity test. Because this test compares phenotypes of reciprocal hemizygotes that each carry a single functional copy of *ebony*, we also examined CHC profiles in *D. americana* and *D. novamexicana* flies with a single functional copy of *ebony*. We found that *D.*

novamexicana flies hemizygous for *ebony* contained a distribution of CHCs biased toward shorter chain hydrocarbons relative to CHCs extracted from *D. americana* flies hemizygous for *ebony*. Prior work found similar profiles of CHCs for wild type *D. americana* and *D. novamexicana* (Bartelt et al., 1986): in both studies, CHCs with a chain length of 25 or fewer were only present in *D. novamexicana*, whereas multiple CHCs with a chain length greater than 30 were only present in *D. americana* (Figure 3-5A). These data suggest that the wild-type *ebony* allele has a dominant effect on CHC profiles.

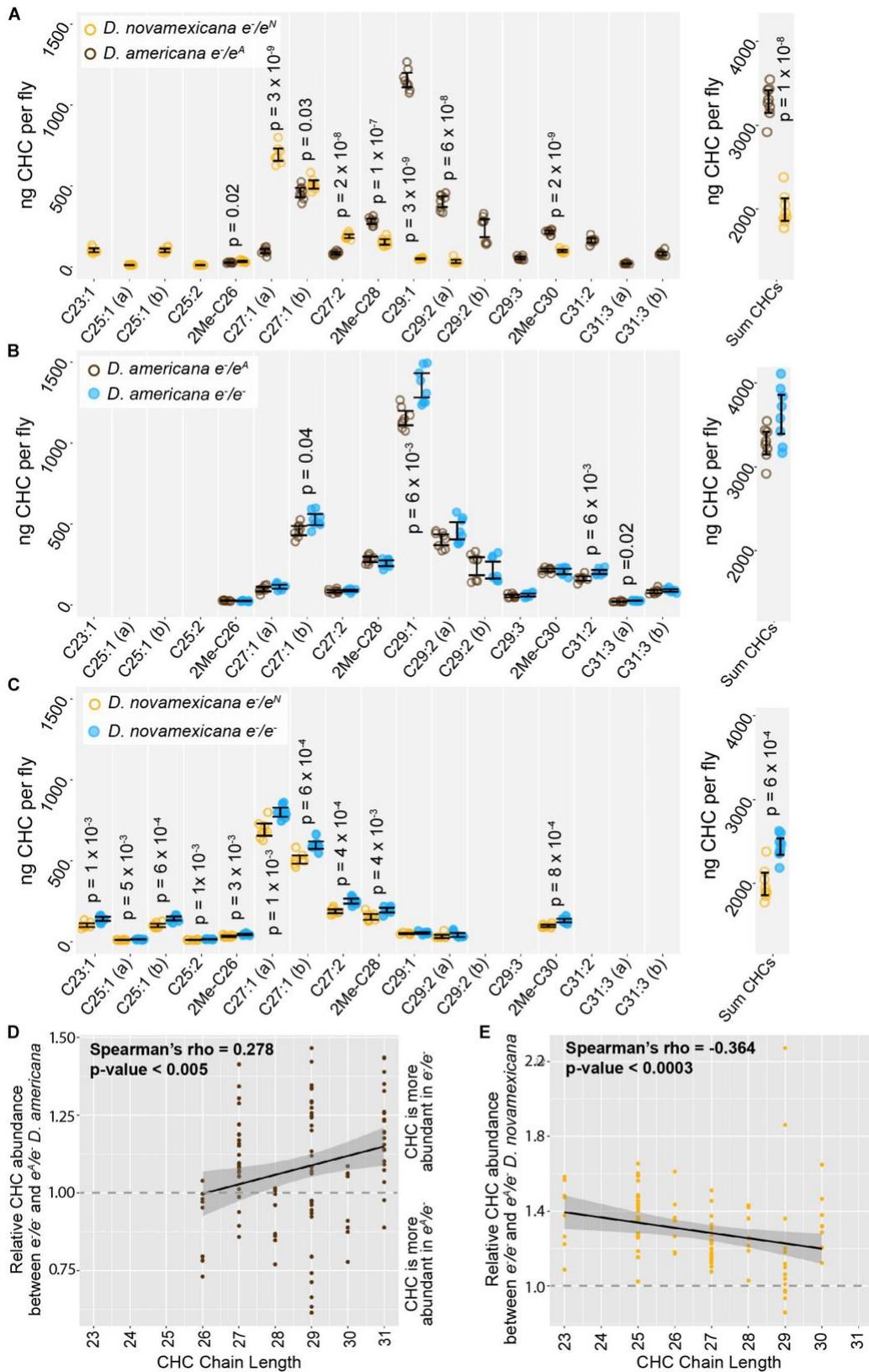


Figure 3-5 Cuticular hydrocarbons (CHCs) are affected by ebony and differ between *D. americana* and *D. novamexicana*. (A–C) Abundance of individual CHC compounds (ng/fly) and

summed CHCs extracted from female flies are plotted for the following genotypes: **(A)** *D. americana* and *D. novamexicana*, each heterozygous for an *ebony* null (e^-) allele, **(B)** *D. americana* heterozygous and homozygous for an *ebony* null allele, **(C)** *D. novamexicana* heterozygous and homozygous for an *ebony* null allele. Eight biological replicates are shown for each genotype, with error bars representing 95% confidence intervals. For each comparison, the p -value from a Welch's t -test with a Benjamini–Hochberg multiple test correction ($\alpha = 0.05$) is shown when a significant difference in abundance was detected for a CHC present in both genotypes being compared. CHCs are shown from left to right with increasing chain length (represented by “C” followed by the chain length) with double-bond and methyl-branched structures indicated by notations after the colon or before the “C”, respectively. For example, C25:1 represents a 25-carbon monoene, C25:2 represents a 25-carbon diene, and 2Me-C28 represents a 28-carbon alkene with a methyl branch at the second carbon. **(D–E)** Abundance of each CHC in *ebony* null mutants relative to flies heterozygous for the *ebony* null allele is plotted by carbon chain length for **(D)** *D. americana* and **(E)** *D. novamexicana*. Black trendlines in panels **(D–E)** show linear regressions, with shaded areas representing the standard error and both Spearman's rho and p -values indicated on each plot.

In order to test whether *ebony* affects CHCs in these species, we compared CHCs extracted from homozygous *ebony* mutants to those extracted from *ebony* heterozygotes. In both species, the loss of *ebony* function had no qualitative effect on which CHCs were produced by either species, but increased the abundance of some CHCs in both *D. americana* and *D. novamexicana* (Figures 3-5B,C). Because *ebony* loss-of-function mutants in *D. melanogaster* were recently shown to preferentially increase the abundance of long chain CHCs (Massey et al., 2019b), we compared relative abundance of individual CHCs between *ebony* null and heterozygous samples and plotted the results against CHC chain length (Figures 3-5D,E). We observed a similar pattern to *D. melanogaster* in *D. americana*, with *ebony* loss-of-function increasing the abundance of longer chain CHCs more strongly (Figure 3-5D). In *D. novamexicana*, we observed the opposite pattern, however: CHCs with shorter chain lengths showed greater increases in abundance in *ebony* null mutants (Figure 3-5E). The reason for this difference in how *ebony* affects CHCs in *D. americana* and *D. novamexicana* remains unclear, but might have to do with the different levels of *tan* expression in these two species (Cooley et al., 2012) given that *tan* was also shown to affect CHC profiles in *D. melanogaster* (Massey et al., 2019b). Specifically, *tan* expression is lower in *D. novamexicana* pupae relative to *D. americana* (Cooley et al., 2012), and *tan* loss-of-function has been shown to preferentially increase the abundance of shorter chain CHCs in *D. melanogaster* (Massey et al., 2019b). Further experiments exploring the mechanisms

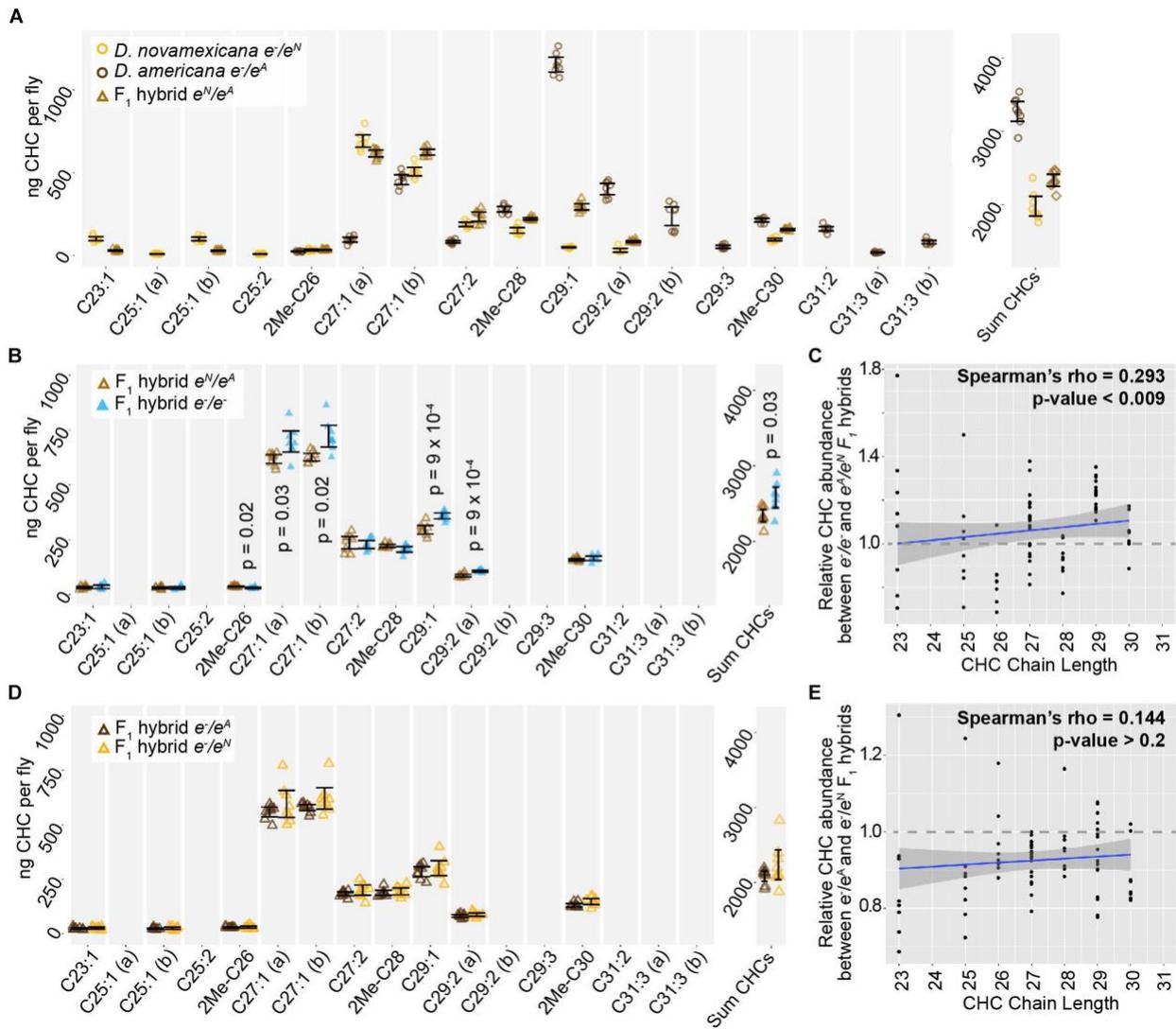


Figure 3-6 *ebony* does not contribute to divergence of CHCs between *D. americana* and *D. novamexicana*. (A) Abundance of individual CHC compounds (ng/fly) and summed CHCs extracted from female flies are plotted for *D. americana* and *D. novamexicana* *ebony* heterozygotes as well as F₁ hybrids heterozygous for wild-type alleles of *ebony*. (B–C) CHCs from F₁ hybrids homozygous for *ebony* null alleles are compared to CHCs from F₁ hybrids with wild-type *D. americana* and *D. novamexicana* *ebony* alleles, showing the absolute abundance of individual and summed CHC compounds (B) as well as the relative abundance of CHCs by carbon chain length (C). In panel (B), *p*-values are shown from a Welch’s *t*-test with a Benjamini–Hochberg multiple test correction ($\alpha = 0.05$) when a significant difference in abundance was detected for a CHC present in both genotypes. (D–E) CHC profiles are plotted for reciprocal F₁ hybrids that differ only by which wild-type *ebony* allele they carry, either *D. americana* (e^A) or *D. novamexicana* (e^N), with absolute abundance of individual and summed CHCs shown in (D) and relative abundance of CHCs by chain length shown in (E). No *p*-values are shown in (D) because no CHCs showed a statistically significant difference in abundance between the two F₁ hybrid genotypes from the reciprocal hemizygosity test (Welch’s *t*-test with Benjamini–Hochberg multiple test correction, $p > 0.05$ for each CHC). In panels (C,E), blue trendlines show linear regressions, with shaded areas representing the standard error and both Spearman’s rho and *p*-values indicated on each plot. In all panels, data from eight replicate flies is shown for each genotype.

underlying CHC production in these species may provide more insight into the contrasting effects of *ebony* on CHCs in *D. americana* and *D. novamexicana*.

We also examined the CHC profiles of female F₁ hybrids produced by crossing *D. americana* females with *D. novamexicana* males. We found that these F₁ hybrid females showed a CHC profile that was distinct from both species, but more similar to *D. novamexicana* (Figure 3-6A): it contained some of the short chain CHCs unique to *D. novamexicana* and none of the long chain CHCs unique to *D. americana* (Figure 3-6A). As seen for both species, eliminating *ebony* function in F₁ hybrids by making them homozygous for *ebony* null alleles caused an increase in abundance of some CHCs but did not alter which CHCs were present (Figure 3-6B). Longer chain CHCs were more likely to show increased abundance than shorter chain CHCs (Figure 3-6C), but this relationship was not as strong as that seen for *D. americana* (Figure 3-5D). To determine whether divergence between the *D. americana* and *D. novamexicana* *ebony* alleles affected CHCs profiles, we compared CHCs extracted from females from the reciprocal hemizyosity test. These flies have only one functional *ebony* allele (*D. americana* or *D. novamexicana*) in the F₁ hybrid genetic background. The CHC profiles from these flies were not significantly different from each other (Figures 3-6D,E), indicating that allelic divergence at *ebony* does not have a detectable effect on CHCs in this species pair.

Conclusions

Identifying the genes responsible for phenotypic differences between species remains a significant challenge for evolutionary biology. This task is especially challenging when a gene contributing to phenotypic divergence is located in a region of the genome inverted between species, which precludes recombination-based mapping. Such is the case for the *ebony* gene in *D. americana* and *D. novamexicana*. Prior work suggested that *ebony* might contribute to differences in overall body color between these two species (Wittkopp et al., 2009; Cooley et al., 2012), but its location in an inversion made it difficult to directly test this hypothesis. In this study, we overcame this hurdle by using CRISPR/Cas9 genome editing to generate null mutants for *ebony* in *D. americana* and *D. novamexicana*, and then using these mutants to perform a reciprocal hemizyosity test (Stern, 2014), which directly compares the effects of the two species'

alleles on pigmentation. We found that divergence at *ebony* does indeed contribute to differences in body color between *D. americana* and *D. novamexicana*.

Characterizing the phenotypes of *D. americana* and *D. novamexicana* *ebony* mutants, as well as flies from the reciprocal hemizyosity test, also identified effects of *ebony* on other phenotypes. For example, we found that differences in the activity of *ebony* alleles between *D. americana* and *D. novamexicana* are responsible for the absence of dark pigmentation seen along the dorsal abdominal midline of *D. novamexicana* but not *D. americana*. This trait has previously been described as derived in *D. novamexicana* (Spicer, 1991); however, we see a similar dorsal midline lightening in at least some lines of *D. lummei* (see Figure 3-1), another member of the virilis group, suggesting that the dorsal midline activity of *ebony* existed prior to the divergence of *D. americana* and *D. novamexicana*. An unexpected change in pupal pigment patterning was also seen in *D. americana* and *D. novamexicana* *ebony* null mutants. Although *ebony* is known to affect pupal case development in *D. melanogaster* (Sherald, 1980), its loss causes a pale white pupa color rather than the dark pigmentation we see in *D. americana* and *D. novamexicana* *ebony* null mutants. Because *ebony* is required for the production of yellow pigments, the dark markings seen in *ebony* mutant pupal cases likely result from expression of an enzyme required for synthesis of dark pigments, such as *tan*. Finally, we found that *ebony* null mutants showed significant changes in the abundance of some CHCs in each species, but divergence of *ebony* did not contribute to differences in the CHC profiles seen between species. These observations illustrate how *cis*-regulatory changes can cause divergence of some, but not all, traits affected by a pleiotropic gene.

Observations reported in this work were made possible by the ability to manipulate the *D. americana* and *D. novamexicana* genomes with CRISPR/Cas9 genome editing. While this technology has great potential for allowing functional hypothesis testing in species that have not historically been considered genetic model systems, this work was not always straightforward. We hope that the detailed descriptions of our genome editing efforts provided in the Materials and “Materials and

Methods” section of this paper will be helpful for other researchers striving to manipulate the genomes of non-model species.

Data Availability Statement

All datasets generated for this study are included in the article/Supplementary Material.

Author Contributions

AL and PW conceived of the experiments and wrote the manuscript. AL, PS, and ZW performed the experiments. HC and PJ provided funding, advice, and oversight. All authors contributed to the article and approved the submitted version.

Funding

This work was funded by the National Institutes of Health (Grant Nos. 1R35GM118073 and 1R01GM089736) awarded to PW; National Science Foundation Graduate Research Fellowship Program (Grant No. DGC 1256260) and National Institute of Health training grant: “Michigan Predoctoral Training in Genetics” (Grant No. T32-GM07544) to AL; and startup funding provided by the Michigan State University AgbioResearch to HC.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

We thank Arnaud Martin (George Washington University) as well as Kathy Vaccarro and other members of Sean Carroll's laboratory (University of Wisconsin) for advice on CRISPR/Cas9 genome editing and *Drosophila* injections, respectively; Hannah McConnell, Aida de la Cruz, and Harmit Malik (Fred Hutchinson Cancer Research Center) for sharing their experience working with the *nanos* promoter in *Drosophila virilis*; and the Bloomington Drosophila Stock Center as well as the National Drosophila Species Stock Center for maintaining and providing fly stocks

References

Ahmed-Braimah, Y.H. and McAllister, B.F. (2012). Rapid evolution of assortative fertilization between recently allopatric species of *Drosophila*. *Int. J. Evol. Biol.* 2012, 285468. doi: 10.1155/2012/285468.

Ahmed-Braimah, Y.H. and Sweigart, A.L. (2015). A single gene causes an interspecific difference in pigmentation in *Drosophila*. *Genetics* 200, 331–42. doi: 10.1534/genetics.115.174920.

Bartelt, R.J., Arnold, M.T., Schaner, A.M., and Jackson, L.L. (1986). Comparative analysis of cuticular hydrocarbons in the *Drosophila virilis* species group. *Comp. Biochem. Physiol. -- Part B Biochem.* 83, 731–42. doi: 10.1016/0305-0491(86)90138-0.

Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.-L. (2013). Highly Efficient Targeted Mutagenesis of *Drosophila* with the CRISPR/Cas9 System. *Cell Rep.* 4, 220–28. doi: 10.1016/j.celrep.2013.06.020.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. <https://www.jstor.org/stable/2346101>.

Caletka, B.C. and McAllister, B.F. (2004). A genealogical view of chromosomal evolution and species delimitation in the *Drosophila virilis* species subgroup. *Mol. Phylogenet. Evol.* 33, 664–70. doi: 10.1016/j.ympev.2004.08.007.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3, e245. doi: 10.1371/journal.pbio.0030245.

Chung, H. and Carroll, S.B. (2015). Wax, sex and the origin of species: Dual roles of insect cuticular hydrocarbons in adaptation and mating. *BioEssays*, 822–30. doi: 10.1002/bies.201500014.

Chung, H., Loehlin, D.W., Dufour, H.D., Vaccarro, K., Millar, J.G., and Carroll, S.B. (2014). A single gene affects both ecological divergence and mate choice in *Drosophila*. *Science* 343, 1148–51. doi: 10.1126/science.1249998.

Cooley, A.M., Shefner, L., McLaughlin, W.N., Stewart, E.E., and Wittkopp, P.J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*. *Evol. Dev.* 14, 317–25. doi: 10.1111/j.1525-142X.2012.00550.x.

Holtzman, S., Miller, D., Eisman, R., Kuwayama, H., Niimi, T., and Kaufman, T. (2010). Transgenic tools for members of the genus *Drosophila* with sequenced genomes. *Fly (Austin)*. 4, 349–62. doi: 10.4161/fly.4.4.13304.

Horn, C. and Wimmer, E.A. (2000). A versatile vector set for animal transgenesis. *Dev. Genes Evol.* 210, 630–37. doi: 10.1007/s004270000110.

Koch, P.B., Behnecke, B., Weigmann-Lenz, M., and Ffrench-Constant, R.H. (2000). Insect pigmentation: Activities of β -alanyldopamine synthase in wing color patterns of wild-type and melanic mutant swallowtail butterfly *Papilio glaucus*. *Pigment Cell Res.* 13, 54–58. doi: 10.1111/j.0893-5785.2000.130811.x.

Kronforst, M.R., Barsh, G.S., Kopp, A., Mallet, J., Monteiro, A., Mullen, S.P., Protas, M., Rosenblum, E.B., Schneider, C.J., and Hoekstra, H.E. (2012). Unraveling the thread of nature's tapestry: the genetics of diversity and convergence in animal pigmentation. *Pigment Cell Melanoma Res.* 25, 411–33. doi: 10.1111/j.1755-148X.2012.01014.x.

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., and Ma, W. (2018). Methodologies for Improving HDR Efficiency. *Front. Genet.* 9, 691. doi: 10.3389/fgene.2018.00691.

Massey, J. H. and Wittkopp, P.J. (2016). The genetic basis of pigmentation differences within and between *Drosophila* species. In *Curr. Top. Dev. Biol.*, 119:27–61. doi: 10.1016/bs.ctdb.2016.03.004.

Massey, Jonathan H., Chung, D., Siwanowicz, I., Stern, D.L., and Wittkopp, P.J. (2019a). The *yellow* gene influences *Drosophila* male mating success through sex comb melanization. *eLife* 8, 1–20. doi: 10.7554/eLife.49388.

Massey, J. H., Akiyama, N., Bien, T., Dreisewerd, K., Wittkopp, P.J., Yew, J.Y., and Takahashi, A. (2019b). Pleiotropic effects of *ebony* and *tan* on pigmentation and cuticular hydrocarbon composition in *Drosophila melanogaster*. *Front. Physiol.* 10, 518. doi: 10.3389/fphys.2019.00518.

Miller, D.F.B., Holtzman, S.L., and Kaufman, T.C. (2002). Customized microinjection glass capillary needles for P-element transformations in *Drosophila melanogaster*. *Biotechniques* 33, 366–75. doi: 10.2144/02332rr03.

Morales-Hojas, R., Vieira, C.P., and Vieira, J. (2008). Inferring the evolutionary history of *Drosophila americana* and *Drosophila novamexicana* using a multilocus approach and the influence of chromosomal rearrangements in single gene analyses. *Mol. Ecol.* 17, 2910–26. doi: 10.1111/j.1365-294X.2008.03796.x.

Nappi, A.J. and Christensen, B.M. (2005). Melanogenesis and associated cytotoxic reactions: Applications to insect innate immunity. *Insect Biochem. Mol. Biol.* 35, 443–59. doi: 10.1016/j.ibmb.2005.01.014.

Patterson, J.T. and Stone, W.S. (1949). The relationship of *novamexicana* to the other members of the *virilis* group. In *Univ. Texas Publ.*, 4920:7–17.

Port, F., Chen, H.M., Lee, T., and Bullock, S.L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 111, E2967-76. doi: 10.1073/pnas.1405500111.

Rebeiz, M. and Williams, T.M. (2017). Using *Drosophila* pigmentation traits to study the mechanisms of cis-regulatory evolution. *Curr. Opin. Insect Sci.* 19, 1–7. doi: 10.1016/j.cois.2016.10.002.

Sherald, A.F. (1980). Sclerotization and coloration of the insect cuticle. *Experientia* 36, 143–46. doi: 10.1007/BF01953696.

Spicer, G.S. (1991). The genetic basis of a species-specific character in the *Drosophila virilis* species group. *Genetics* 128, 331–37. <http://www.ncbi.nlm.nih.gov/pubmed/2071018>.

- Stern, D.L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizyosity test. *Trends Genet.* 30, 547–54. doi: 10.1016/j.tig.2014.09.006.
- Stern, D.L., Crocker, J., Ding, Y., Frankel, N., Kappes, G., Kim, E., Kuzmickas, R., Lemire, A., Mast, J.D., and Picard, S. (2017). Genetic and transgenic reagents for *Drosophila simulans*, *D. mauritiana*, *D. yakuba*, *D. santomea*, and *D. virilis*. *G3 (Bethesda)*. 7, 1339–47. doi: 10.1534/g3.116.038885.
- Suh, J. and Jackson, F.R. (2007). *Drosophila ebony* activity is required in glia for the circadian regulation of locomotor activity. *Neuron* 55, 435–47. doi: 10.1016/j.neuron.2007.06.038.
- Takahashi, A. (2013). Pigmentation and behavior: potential association through pleiotropic genes in *Drosophila*. *Genes Genet. Syst.* 88, 165–74. <http://www.ncbi.nlm.nih.gov/pubmed/24025245>.
- Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., Kaufman, T.C., Calvi, B.R., and FlyBase Consortium. (2019). FlyBase 2.0: the next generation. *Nucleic Acids Res.* 47, D759–65. doi: 10.1093/nar/gky1003.
- True, J.R. (2003). Insect melanism: the molecules matter. *Trends Ecol. Evol.* 18, 640–47. doi: 10.1016/j.tree.2003.09.006.
- True, J.R., Yeh, S.D., Hovemann, B.T., Kemme, T., Meinertzhagen, I. a, Edwards, T.N., Liou, S.R., Han, Q., and Li, J. (2005). *Drosophila tan* encodes a novel hydrolase required in pigmentation and vision. *PLoS Genet.* 1, e63. doi: 10.1371/journal.pgen.0010063.
- Wittkopp, P. J., Stewart, E.E., Arnold, L.L., Neidert, A.H., Haerum, B.K., Thompson, E.M., Akhras, S., Smith-Winberry, G., and Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science*. 326, 540–44. doi: 10.1126/science.1176980.
- Wittkopp, P J, Smith-Winberry, G., Arnold, L.L., Thompson, E.M., Cooley, a M., Yuan, D.C., Song, Q., and McAllister, B.F. (2011). Local adaptation for body color in *Drosophila americana*. *Heredity (Edinb)*. 106, 592–602. doi: 10.1038/hdy.2010.90.
- Wittkopp, Patricia J, and Beldade, P. (2009). Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Semin. Cell Dev. Biol.* 20, 65–71. doi: 10.1016/j.semcd.2008.10.002.
- Wittkopp, Patricia J, Carroll, S.B., and Kopp, A. (2003a). Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet.* 19, 495–504. doi: 10.1016/S0168-9525(03)00194-X.

Wittkopp, Patricia J, True, J.R., and Carroll, S.B. (2002). Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* 129, 1849–58. <http://www.ncbi.nlm.nih.gov/pubmed/11934851>.

Wittkopp, Patricia J, Williams, B.L., Selegue, J.E., and Carroll, S.B. (2003b). *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1808–13. doi: 10.1073/pnas.0336368100.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M. V, and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419. doi: 10.1093/molbev/msg140.

Chapter 4

microRNAs Are Necessary Components of the Genetic Architecture Underlying Adult Cuticle Pigmentation in *Drosophila melanogaster*.

Abstract

The genetic architecture encoding all phenotypes incorporates multiple levels of regulation to ensure gene expression is tightly controlled. While most research into the genetic basis of phenotypes and the regulation of gene expression has focused on the process of transcription, post-transcriptional regulation is increasingly understood to be an important component of development in metazoans. One critical mechanism of post-transcriptional regulation is microRNA-induced silencing or degradation of messenger RNAs, preventing the production of proteins from targeted transcripts. However, the roles of microRNAs within developmental networks are poorly understood. In this study, we examined the effects of microRNAs on cuticular pigmentation in adult *Drosophila melanogaster*, a system that has been extensively studied as a model of genetic regulation and the evolution of gene expression. We overexpressed a collection of 166 miRNAs in the dorsal midline of developing flies and found that 48 were sufficient to affect pigmentation. We further investigated the endogenous effects of 41 miRNAs on pigmentation by competitively inhibiting them in the same tissue, finding that 22 were necessary for the development of wildtype pigment patterns. We then identified candidate miRNA-target interactions through computational predictions. Functional testing of a subset of potential miR-8 targets revealed evidence of coordinated regulation of multiple genes with similar effects on pigmentation, suggesting a possible

mechanism for this miRNA's action within the genetic network underlying *D. melanogaster* pigmentation.

Introduction

Proper development of a multicellular organism requires strict control of gene expression, with genes expressed in the necessary time, place, and environmental context. This expression is controlled first by transcriptional regulation, in which enhancers and promoters interact with transcription factors to determine when, where, and how much RNA is transcribed from a gene. After RNA transcripts are made, post-transcriptional regulation further impacts the expression of gene products by altering the stability, splicing, capping, polyadenylation, and translational efficiency of RNAs (Halbeisen et al., 2008). microRNAs (miRNAs) are small, non-coding RNAs that act as important post-transcriptional regulators by guiding the RNA-induced silencing complex (RISC) to the 3' UTRs of messenger RNAs (mRNAs) and preventing their translation into protein (Bartel, 2018). Despite their key role in regulating gene expression, many questions remain about how post-transcriptional regulation by miRNAs impacts the gene regulatory networks that control development.

Individual miRNAs were once thought to have little effect on phenotypes because early studies of miRNAs in *Caenorhabditis elegans* showed that loss of a single miRNA's function often had no discernable impacts on the phenotypes assayed (Miska et al., 2007). More recent studies in *Drosophila*, however, have shown that loss of miRNA function can have strong effects on many phenotypes, including viability, lifespan, fertility, and various morphological, physiological and behavioral phenotypes (Chen et al., 2014; Fulga et al., 2015; Garaulet et al., 2020; Picao-Osorio et al., 2015, 2017; Verma and Cohen, 2015). Natural variation in miRNA expression has also been shown to affect the pattern of leg trichomes in *Drosophila* (Arif et al., 2013), suggesting that miRNAs can also contribute to the evolution of development. Yet it remains unclear how miRNAs affect the development of complex traits. For example, how many miRNAs

typically affect a trait? How do these miRNAs work with other genes (e.g., transcription factors, effector genes) involved in the development of the same trait? And to what extent do individual miRNAs impact a given trait? Here, we approach these questions by systematically studying the role of miRNAs in the development of pigmentation in adult *Drosophila melanogaster*, a trait that has served for decades as a model system in studies of genotype-phenotype relationships (Massey and Wittkopp, 2016a).

Drosophila pigmentation has been used as a model system to understand the regulation of gene expression, developmental processes, and mechanisms of phenotypic evolution (Massey and Wittkopp, 2016a; Rebeiz and Williams, 2017). Many of the genes required for the synthesis of pigments that make up adult pigment patterns are well-characterized (Massey and Wittkopp, 2016a), as are many of the transcription factors that regulate (either directly or indirectly) the expression of these genes (Kalay et al., 2016; Rogers et al., 2014). Prior work has shown that at least one miRNA, *miR-8*, affects pigment patterning in adult flies: when *miR-8* is mutated or competitively inhibited, dark melanin is reduced in the abdomen (Kennell et al., 2012). To more systematically search for miRNAs that impact pigmentation development, we overexpressed 166 miRNAs along the dorsal midline of developing flies and found that more than one quarter of these miRNAs (48/166) were sufficient to cause a visible change in pigmentation in the adult cuticle. We competitively inhibited 41 miRNAs chosen from among those whose overexpression affected pigmentation, and found that 22 of these were also necessary for the development of normal pigmentation. 16 of these miRNAs showed opposite phenotypes in response to overexpression and inhibition of the miRNA, suggesting that the miRNA plays a critical role in pigmentation development. Surprisingly, the magnitude of effects on pigmentation caused by these miRNAs were similar to the magnitude of effects caused by the knockdown of many transcription factors shown to affect pigmentation (Kalay et al., 2016; Rogers et al., 2014).

We further investigated predicted targets of *miR-8*, and found that this miRNA appears to promote the production of dark pigmentation in the posterior female

abdomen by coordinately repressing a suite of genes that function in the development of light yellow pigments rather than dark melanins. In addition, we identify several promising miRNA-target pairs for further investigation into their role in the genetic network that regulates pigmentation development in *Drosophila*. Together, these experiments reveal a previously under-appreciated role of post-transcriptional regulation in adult *D. melanogaster* pigmentation, making this system a promising model system for the study of miRNA regulation within a genetic network and its effects on phenotypes.

Results and Discussion

To identify miRNAs that regulate the development and synthesis of pigments in *D. melanogaster*, we used the UAS/Gal4 system (Brand and Perrimon, 1993) to overexpress 166 miRNAs in developing flies and assessed their effects on pigmentation. This set of miRNAs includes 85.7% of all confidently annotated miRNAs in *D. melanogaster* (Kozomara et al., 2019), and 97% of the available UAS-miRNA lines from the collection described in Schertel et al. (Schertel et al., 2012). We chose to use *pannier*-Gal4 (*pnr*-Gal4) to overexpress miRNAs, since this driver expresses Gal4 in the dorsal midline of pupae during the stages in which pigmentation develops (Kalay et al., 2016; Rogers et al., 2014; Wittkopp et al., 2002). Because the expression pattern of *pnr*-Gal4 forms a clearly-defined stripe along the anterior-posterior body axis, lateral regions of each fly could be used as an internal control to compare pigmentation between regions of dorsal abdominal cuticle with and without miRNA expression within the same animal (Figure 4-1). This same Gal4 driver was later used to express miRNA sponges to competitively inhibit 41 miRNAs to determine whether these miRNAs also affect pigmentation endogenously (Fulga et al., 2015). Pigmentation of flies carrying both the UAS and Gal4 constructs was scored by visually comparing pigmentation in the dorsal midline region of the abdomen to the more lateral abdominal regions of the same animal. While we initially planned to also compare flies inheriting *pnr*-Gal4 to siblings inheriting the TM6B balancer present in this Gal4 line, we found that flies

carrying TM6B were unusually darkly pigmented. This is likely due to the *ebony* mutation on this balancer chromosome, which, while classically described as recessive, appears to display incomplete dominance. For each fly phenotyped, the pigmentation of the dorsal midline where the miRNA was mis-expressed was scored as either “unaffected” (i.e., dorsal midline pigmentation was not noticeably lighter or darker than the lateral regions for any segments) or either “lightened” or “darkened” if the dorsal midline was noticeably lighter or darker than lateral regions in any segment. We found that the vast majority of flies phenotyped had extremely consistent phenotypes, making those with darker or lighter pigmentation easy to recognize. In some cases, the dorsal midline was darker than the lateral regions in one body segment, but lighter than the lateral regions in another body segment, so the phenotypes for the different body segments were recorded individually. Pigmentation was scored for an average of 10 female and 9 male flies per cross. Noticeably lightened or darkened pigmentation phenotypes were often observed in a subset of siblings within a single cross, and only genotypes with consistent effects observed in at least 50% of flies are included in the counts below. Observations for each individual fly with each genotype are provided as Supplemental Table S4-1.

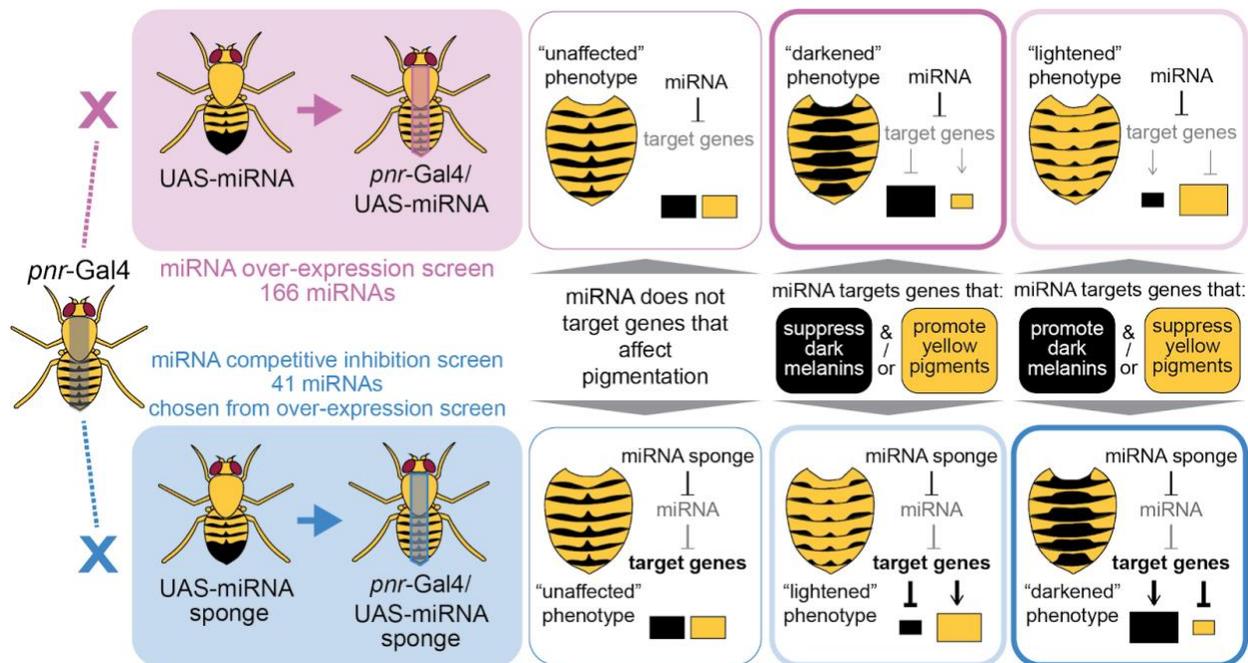


Figure 4-1 Schematic of miRNA overexpression and competitive inhibition screens and scoring parameters. Virgin female flies carrying *pnr-Gal4* were crossed to males carrying UAS-miRNA for the overexpression screen (shown in pink), or to males carrying UAS-miRNA sponge for the competitive

inhibition screen (shown in blue). Offspring from these crosses have their respective UAS-driven transgenes expressed down the dorsal midline of the fly, either overexpressing or inhibiting the miRNA under investigation. These offspring were classified as “unaffected”, “darkened”, or “lightened” based on the pigmentation phenotype in the *pnr*-Gal4 expression domain. Cartoon schematics represent hypothetical genetic relationships mediating the effects of miRNAs on pigmentation. Downward-facing arrows and nail heads represent positive and negative regulation, respectively. Target genes that are expected to have reduced expression in response to miRNA manipulation are shown in gray, while those that are expected to have increased expression (de-repression) are shown in bold. Black and yellow boxes represent the production of dark melanins and light yellow pigments, respectively. Larger or smaller black or yellow boxes represent increased or decreased production of pigments in response to miRNA overexpression or competitive inhibition.

Overexpression of 48 individual miRNAs is sufficient to lighten or darken pigmentation in one or more body segments.

Of the 166 miRNAs tested, the ratio of progeny inheriting both Gal4 and UAS-miRNA versus those inheriting a balancer in place of either transgene was significantly less than the expected mendelian ratio for 19 miRNAs, (one-sided binomial test, $p < .05$) suggesting that overexpression of the miRNA along the dorsal midline with the *pnr*-Gal4 driver reduced viability (Table 4-1). Of these 19 crosses with reduced viability, 13 produced 2 or fewer progeny carrying both Gal4 and UAS-miRNA, which we defined as a “lethal” phenotype. Of the 153 miRNAs that could be overexpressed non-lethally with *pnr*-Gal4, we found 105 did not cause a noticeable change in pigmentation at a penetrance of 50% or higher in any body segment. Of the remaining 48 miRNAs, only counting phenotypes with a penetrance of 50% or higher, overexpression of 25 caused lighter pigmentation in one or more body segments, 29 caused darker pigmentation in one or more body segments, and 6 caused lighter pigmentation in some segments but darker in others (Figure 4-2, Table 4-1). Of the miRNAs that affected pigmentation, the most common phenotypes were darkening or lightening of the posterior female abdomen, generally restricted to the A6 abdominal segment, but occasionally extending into A5 (Figure 4-2). Much rarer were miRNAs that affected pigmentation in either anterior abdominal segments (A1-A4) or the thorax. Only one of the miRNAs tested visibly affected male posterior abdomen pigmentation, which is likely because male A5 and A6 are fully melanized in wild type male animals, meaning any miRNAs that promote the production of dark pigments in the posterior male abdomen were unlikely to

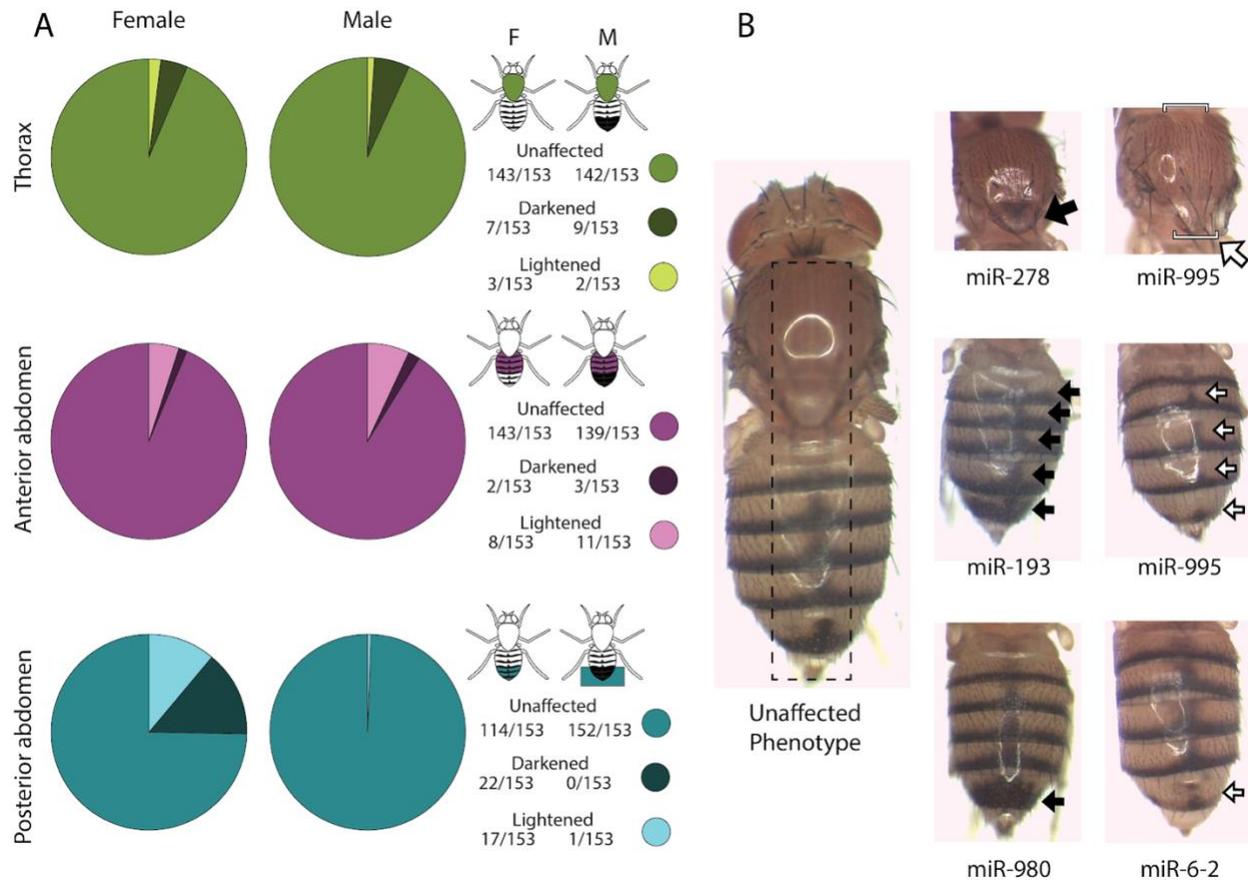


Figure 4- 2: Pigmentation phenotypes resulting from overexpressing 153 miRNAs. (A) Pie charts depict the counts of miRNAs that were sufficient to lighten or darken pigmentation, as well as those that did not affect pigmentation for male and female flies. Each row represents a broad category of body segments along the anterior-posterior body axis: thorax, anterior abdomen (A1-A4), posterior abdomen(A5-A6) **(B)** Representative examples of flies with “lightened”, “darkened”, and “unaffected” phenotypes in each anatomical category. Black arrows and white arrows represent areas of increased and decreased pigmentation, respectively. The “unaffected phenotype” image represents the consistent phenotype observed in the vast majority of flies without balancer phenotypes.

be identified and any that lighten pigmentation would need to do so quite a bit to be readily detectable.

Surprisingly, the phenotypes resulting from overexpression of many miRNAs were qualitatively similar in effect to phenotypes observed in published transcription factor RNAi screens using the same pnr-Gal4 driver. For instance, many of the phenotypes observed in two separate surveys of transcription factors affecting *D. melanogaster* abdominal pigmentation were limited to or more prominent in the A5-A6 segments of female flies (Kalay et al., 2016; Rogers et al., 2014). Interestingly, the female A5-A6 region displays wide phenotypic variation both within and between

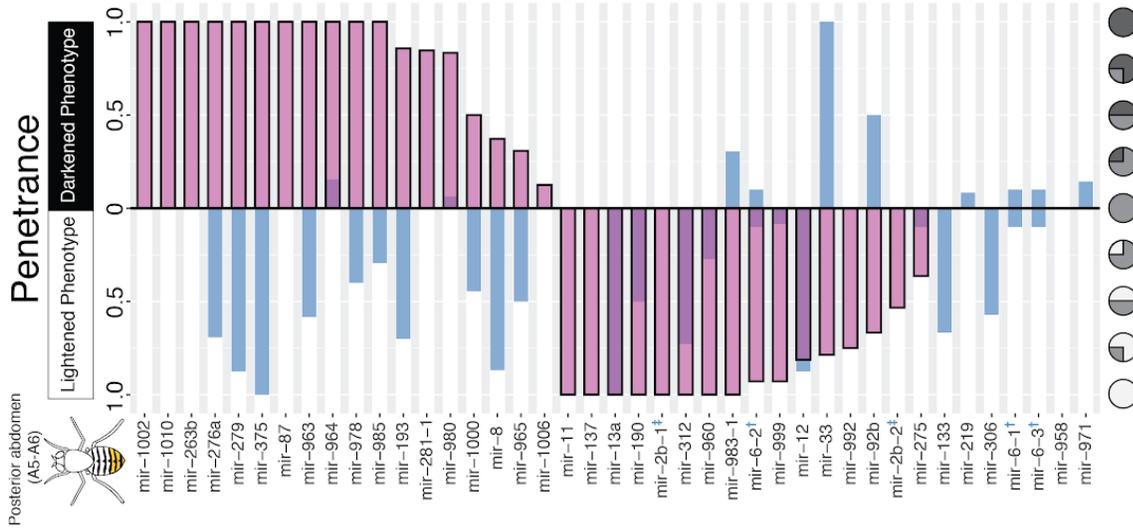
et al., 2007; Salomone et al., 2013; Yassin et al., 2016). In addition, the only miRNA with a demonstrated role in *D. melanogaster* pigmentation, miR-8, is limited in effect to the posterior female abdomen (Kennell et al., 2012). The results of this screen add to a large body of evidence suggesting that the female A5-A6 pigmentation in *Drosophila* is particularly sensitive to genetic and environmental perturbations. Surprisingly, we did not find a strongly penetrant effect on pigmentation when overexpressing miR-8 in this screen. Because previous experiments demonstrating the effects of miR-8 on pigmentation all employed miR-8 loss-of-function (genetic mutation and competitive inhibition), it is possible that miR-8 is natively expressed at a high enough level in the developing cuticle to repress its target genes to a sufficient extent that additional miR-8 expression does not exert any further effects.

Native expression of 22 miRNAs is required for development of normal abdominal pigmentation

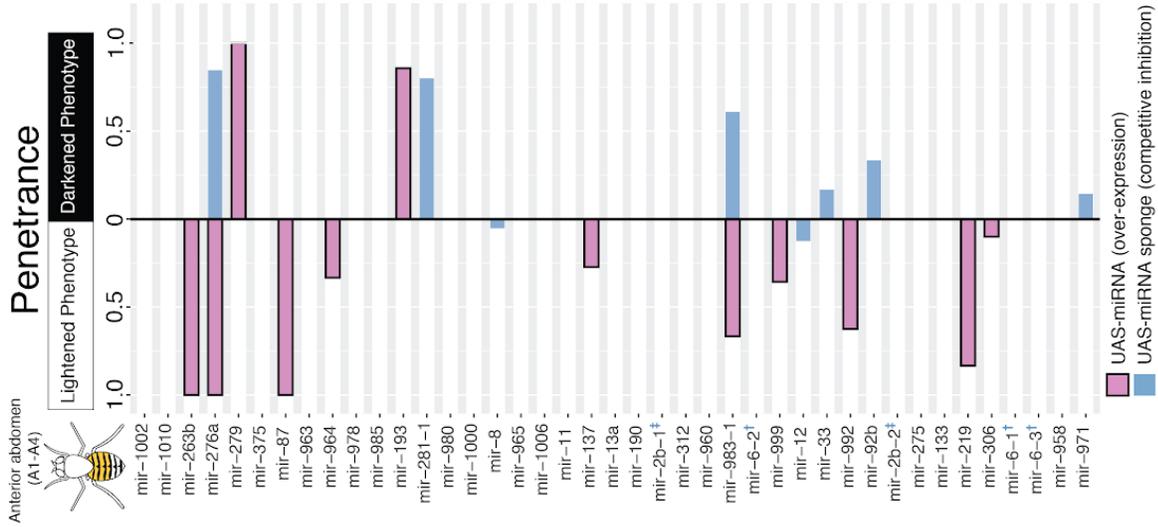
While this overexpression experiment demonstrates which miRNAs are sufficient to affect pigmentation, the miRNAs identified in this screen may or may not be expressed in the correct tissues and developmental stages to affect pigmentation. We therefore followed this experiment by using the same *pnr*-Gal4 driver to competitively inhibit a subset of miRNAs that were identified as sufficient to alter pigmentation in the overexpression screen, as well as miR-8 in an effort to repeat the published phenotype of lightened pigmentation with loss of miR-8 function. We used miRNA “sponges”, which are comprised of RFP-labeled transgenes under UAS-control with 3' UTRs containing 20 recognition sites for a specific miRNA (or 2-3 miRNAs with identical recognition sites in some cases). Ectopic expression of these miRNA recognition sites can out-compete native miRNA targets for miRNA binding, effectively suppressing the effects of the targeted miRNA in the dorsal midline during pigmentation development (Fulga et al 2015). Because we suspected that competitive inhibition might cause more subtle effects on pigmentation than overexpression, depending on the native expression level and effect size of each miRNA, we set each sponge cross in duplicate at two temperatures that cause darker (18°C) or lighter (28°C) overall body pigmentation

(Gibert et al., 2004). We reasoned that lightening phenotypes might be easier to distinguish on more darkly pigmented flies

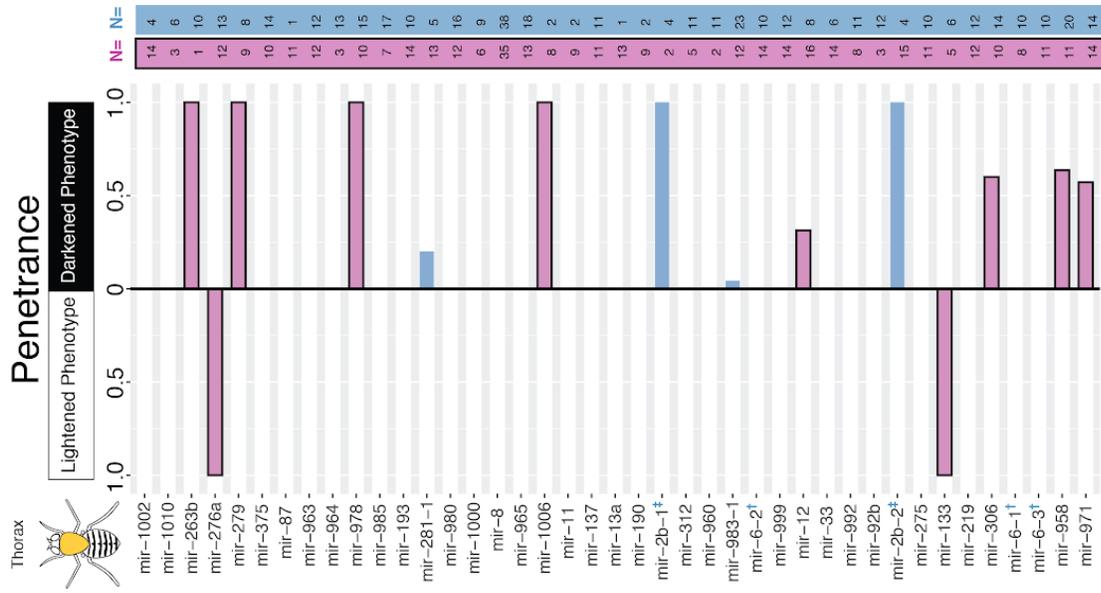
Posterior abdomen
(A5-A6)



Anterior abdomen
(A1-A4)



Thorax



miRNA	N
mir-1002	14
mir-1010	4
mir-263b	6
mir-276a	1
mir-279	10
mir-375	12
mir-87	13
mir-963	9
mir-964	8
mir-978	10
mir-985	14
mir-193	1
mir-281-1	12
mir-980	16
mir-1000	12
mir-8	6
mir-965	35
mir-1006	38
mir-11	13
mir-137	8
mir-13a	2
mir-190	14
mir-2b-1*	10
mir-312	9
mir-960	2
mir-983-1	11
mir-6-2†	12
mir-999	23
mir-12	14
mir-33	10
mir-992	14
mir-92b	8
mir-2b-2*	16
mir-275	8
mir-133	14
mir-219	6
mir-306	15
mir-6-1†	4
mir-6-3†	11
mir-958	10
mir-971	8
mir-958	11
mir-971	20
mir-958	14
mir-971	14

Figure 4-3. Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen. From left to right, the plots represent penetrance of phenotypes observed in the posterior abdomen, anterior abdomen, and thorax of female flies. Penetrance of phenotypes range from 0 (no flies observed display plotted phenotype) to 1 (all flies observed display plotted phenotype). Penetrance of lightening phenotypes are plotted extending to the left from the “0” axis, while penetrance of darkening phenotypes are plotted extending to the right. Competitive inhibition data is represented by blue bars without borders, while over-expression data is represented by pink bars with black borders. Numbers of flies phenotyped for each penetrance value are listed on the right of the figure under “N=”, and are color coded in the same manner as penetrance data. All competitive inhibition data displayed was collected from crosses set at 28°C. Data from competitive inhibition crosses set at 18°C as well as male data are shown in the same format as this figure in Supplemental Figure S4-1, Supplemental Table S4-2, and Supplemental Figure S4-3. † - miR-6-1, miR-6-2, and miR-6-3 share a seed sequence, and thus were simultaneously inhibited by a single miRNA sponge. The competitive inhibition data for these 3 lines are duplicated to display alongside the overexpression data of the corresponding individual miRNAs. ‡ - miR-2b-1 and miR-2b-2 also share a seed and sponge construct, and are displayed in the same manner as the miR-6 family. Summary data from all screen crosses is available in Supplemental Table S4-3

reared at 18°C, while darkening phenotypes might be easier to distinguish on more lightly pigmented flies reared at 28°C.

Using the same scoring parameters as in the overexpression screen, we found that of these 41 miRNAs, 22 caused noticeable differences in pigmentation in the dorsal midline in at least one body segment and one rearing temperature at a penetrance over 50% (Figure 4-3, Table 4-2). While a frequency of over 50% of competitively inhibited miRNAs causing changes in pigmentation was surprisingly high, we note that the miRNA sponge lines we chose to assay were selected from among miRNAs that had already shown pigmentation phenotypes in our overexpression screen. Furthermore, large numbers of miRNAs were found to affect adult flight muscle development in a competitive inhibition screen using the same miRNA sponge collection, with 24% (14/58) of miRNAs having a detectable effect on this phenotype (Fulga et al., 2015). In addition, a survey of miRNA effects on self-righting behavior in *D. melanogaster* larvae found that 41% of miRNA mutant line assayed significantly affected this phenotype (Picao-Osorio et al., 2017). Together, these data suggest that many phenotypes may rely on the action of many miRNAs in their development.

In 15 of the 22 miRNA sponge lines affecting pigmentation, overexpressing the same miRNA caused the opposite phenotype (i.e. darkening versus lightening), suggesting that these miRNAs are both necessary for the development of normal

pigmentation and sufficient to alter pigmentation (Figure 4-3, Table 4-2, Supplemental Figure S4-1, Supplemental Figure S4-2, Supplemental Figure S4-3, Supplemental Table S4-3). Representative images of the phenotypes caused by both overexpression and competitive inhibition of 4 of these miRNAs (miR-33, miR-279, miR-92b, and miR-276a) are shown in Figure 4-4. Of the 15 miRNAs that were both sufficient and necessary to produce pigmentation phenotypes, 6 showed a phenotype in one body segment when overexpressed, but the opposite phenotype was observed upon competitive inhibition only in a different body segment. For instance, overexpressing miR-306 caused darkening in the thorax (scutellum), while competitively inhibiting miR-306 caused lightening of the female posterior abdomen but not the scutellum (Figure 4-3, Table 4-2). These miRNAs may natively act on pigmentation development in the body segment(s) where competitive inhibition causes a change in pigmentation, but, as we

	Overexpression phenotype						Competitive inhibition phenotype					
	Female			Male			Female			Male		
	PA	AA	T	PA	AA	T	PA	AA	T	PA	AA	T
miR-985	D						L*(29%)					
miR-965	D (31%)						L*					
miR-983-1	L	L					L†	D*				
miR-8	D (37%)						L*†				L†	
miR-958			D			D						
miR-275						D						
miR-263b	D	L	D		L	D	L†					
miR-279	D	D	D		D	D	L*†				L†	
miR-13a	L						L*				L†	
miR-92b	L						D*					
miR-133			L			L	L*					
miR-999	L											
miR-992	L	L										
miR-980	D										L*	
miR-978	D		D		D		L*(40%)					
miR-964	D				L							
miR-963	D				D		L*				L†	
miR-960	L											
miR-375	D						L*				L†	
miR-6-1												
miR-6-2	L				L							
miR-6-3												
miR-137	L											
miR-190	L				L		L*					
miR-193	D	D					L*					
miR-219		L			L		D†					
miR-276a	D	L	L	L	L	L	L*	D*				
miR-281-1	D							D*				
miR-12	L						L*				L*†	
miR-11	L				L							
miR-1006			D			D						
miR-1000	D										L†	
miR-312	L					D	L*					
miR-306			D			D	L*					
miR-87	D	L			L							
miR-33	L						D*†					
miR-1002	D											
miR-1010	D											
miR-2b-1	L									D*		D*
miR-2b-2	L									D*		D*
miR-971			D			D	*	*	*	*	*	*

Table 4-2. Summary of phenotypic effects of overexpression and competitive inhibition of all microRNAs included in competitive inhibition screen (Previous page). Column headings: “PA” - posterior abdomen, “AA” - anterior abdomen, “T” - thorax, Abbreviations for results: “L” - dorsal midline lightened, “D” - dorsal midline darkened. Competitive inhibition phenotypes marked with * were observed in crosses set at 28°C, while those marked with † were from crosses set at 18°C. For miR-971, competitive inhibition data is only represented from crosses set at 28°C because the cross set at 18°C failed. All fields marked with “L” or “D” represent crosses where either lightening or darkening phenotypes were observed at a penetrance $\geq 50\%$, except where penetrance values are listed in parentheses. miRNAs that cause opposite phenotypes when overexpressed versus competitively inhibited are shown in bold. miR-6-1, miR-6-2, and miR-6-3 share a seed sequence, as do miR-2b-1 and miR-2b-2. For these miRNAs competitive inhibition data was performed with a single sponge targeting all miRNAs with a common seed, and data obtained with these sponge lines are reproduced in each row that contains overexpression data from the individual miRNAs.

observed with miR-8, may normally be expressed highly expressed enough that overexpression did not alter pigmentation in the miRNA’s native context. Alternatively, it is possible that the miRNA sponge could be interfering with other miRNAs with similar seed sequences to the targeted miRNA, leading to the observed pigmentation changes. However, we note that the sponge lines used in this study have been assessed for specificity and off-target effects, and any miRNA sponge transgenic lines with observed off-target effects were reportedly removed from the miRNA sponge line collection (Fulga et al., 2015). Beyond the 15 miRNAs we identified as necessary and sufficient to affect pigmentation at a penetrance of 50% or higher, an additional 4 miRNAs (miR-8, miR-965, miR-985, miR-978, see Table 4-2) showed opposite phenotypes in overexpression vs competitive inhibition, but were excluded from the main list because either the overexpression or competitive inhibition phenotypes were observed at a penetrance below 50%, suggesting further experiments may be needed to determine their role in pigmentation development.

In the process of phenotyping pnr-Gal4/UAS-miRNA sponge flies, we noticed that, while many lines showed pigmentation differences when the sponge line was expressed down the dorsal midline, the difference between the dorsal midline and lateral dorsal cuticle was rarely as distinct as in the miRNA overexpression crosses. However, we were confident of the phenotypes since these flies differed noticeably from those over-expressing the “scrambled” control sponge, which was designed to be identical to the miRNA sponge lines while not targeting any miRNA in the *D. melanogaster* genome, thus serving as a negative control. For example, flies expressing

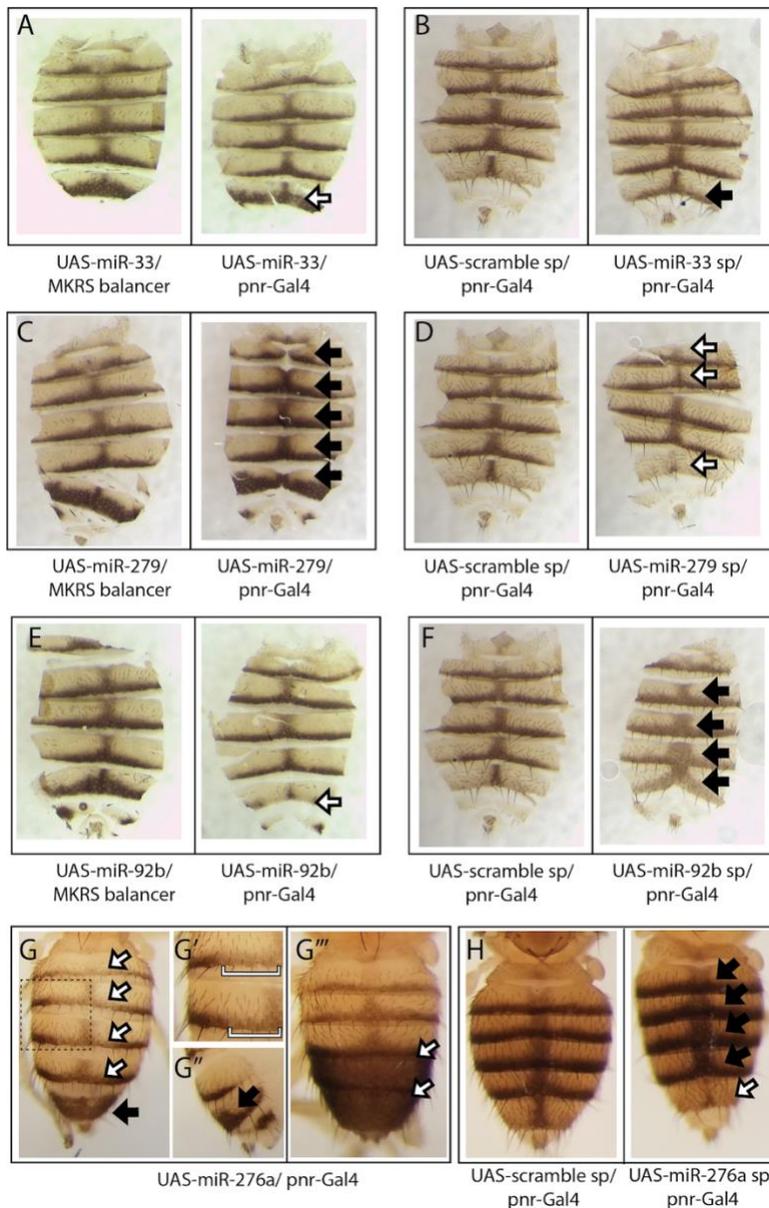


Figure 4-4. Images demonstrating miRNAs that are both sufficient to alter abdominal pigmentation and necessary for normal pigmentation development. In all images, white arrows or brackets indicate areas of lightened pigmentation in the pnr-Gal4 expression domain. while black arrows represent areas of darkened pigmentation in the pnr-Gal4 expression domain. **A-F**: dissected and mounted dorsal abdominal cuticles of female flies comparing flies carrying transgenes to either overexpress (**A,C, E**) or competitively inhibit (**B, D, F**) individual miRNAs compared to matched control flies (left panels of **A-F**). Images depict results from manipulating miR-33 (**A,B**), miR-279 (**C,D**), and miR-92b (**E,F**). **G-H**: images of whole fly abdomens carrying transgenes to overexpress (**G-G''**) or competitively inhibit (**H**) miR-276a. **G**: dorsal view of abdomen of female fly over-expressing miR-276a along the dorsal midline. **G'**: enlarged view of the region outlined in dashed lines in panel G, with area of lightened pigmentation marked by white brackets. **G''**: right lateral view of abdominal segments A5 and A6 from specimen pictured in panels G and G'. Black arrow indicates sharp increase in melanization at the border of the pnr-Gal4 expression domain. **G'''**: dorsal view of abdomen of male fly overexpressing miR-276a with white arrows indicating reduced melanization of segments A5 and A6. **H**: images of whole fly abdomens carrying transgenes to express miRNA sponge constructs targeting either a scrambled region as a negative control (left) or targeting miR-276a.

the sponge transgenes targeting miR-92b and miR-276a both appeared to have darker pigmentation along the entire width of the posterior edge of affected tergites, rather than being confined entirely to the pnr-Gal4 expression domain along the dorsal midline (Figure 4-4F,H). Since we did not observe this phenomenon in the pnr-Gal4/UAS-miRNA flies using the same pnr-Gal4 line, we wondered whether this may be due to “leaky” expression of the sponge transgenes in the absence of Gal4. We confirmed this suspicion by observing the UAS-miRNA sponge lines from the original stock vials, which do not contain a Gal4

transgene, under a fluorescence stereo microscope, where we saw clear RFP signal across the bodies of many of these flies (data not shown). The only RFP reported in these flies' genotype is the coding sequence of the miRNA sponge transcripts themselves, suggesting that these sponge transgenes are expressed in the absence of Gal4 activation, though this expression was noticeably weaker than the RFP signal along the dorsal midline of the flies that inherited both pnr-Gal4 and the UAS-sponge transgenes. We suggest that this leaky miRNA sponge expression should be taken into account when designing and interpreting experiments using these miRNA sponge transgenic lines.

During the course of the screen experiments described in this manuscript, another group independently identified miR-33 as a regulator of *D. melanogaster* pigmentation by observing darkened pigmentation in the A5 and A6 segments of miR-33 mutants relative to wildtype individuals (J. Kennell, unpublished personal communication). Their results agree with the observations we report in this study. We found that overexpressing miR-33 caused a mild but noticeable and penetrant reduction in the dark pigment stripe along the posterior edge of the A6 segment in females, while competitively inhibiting miR-33 reliably caused an increase in dark melanization in female A6 (Figure 4-3, Figure 4-4A-B).

One noteworthy miRNA, miR-276a, showed a particularly striking set of phenotypes in these screen experiments. Interestingly, overexpression of this miRNA caused distinctly opposite phenotypes in the anterior abdomen vs the posterior abdomen of female flies. In abdominal segments anterior to A6, pigmentation was much lighter in the pnr-Gal4 expression domain relative to the dorsal cuticle outside this region, with a sharp boundary between the affected and unaffected portions of the cuticle (Figure 4-4G-G'). In contrast, the A6 segment of female abdomens was almost fully melanized within the boundaries of the pnr-Gal4 expression domain, again with a sharp transition between affected and unaffected cuticle (Figure 4-4G,G'') Furthermore, miR-276a was the only miRNA out of 153 overexpression genotypes to show a clear phenotype in the male posterior abdomen, where pigmentation was noticeably lightened

in the *pnr-Gal4* expression domain relative to the lateral dorsal cuticle (Figure 4-4G'''). The opposite effect of miRNA expression in the posterior abdomen depending on sex was unique to miR-276a. Excitingly, the sponge transgene targeting miR-276a caused a strong lightening effect in the female A6 segment, the pigmentation in anterior abdomen segments of the same flies appeared darker than scrambled sponge controls (Figure 4-4H). Competitive inhibition of miR-276a did not cause a detectable phenotype in the fully melanized male A5 and A6 segments, which is to be expected if, as suggested by the overexpression phenotype, this miRNA represses the formation of dark pigments in male A5 and A6. These results suggest that miR-276a may be an important regulator of sex-specific pigmentation in the posterior abdomen, a trait that has been the subject of a large body of research on the development and evolution of sexually dimorphic phenotypes (Gompel and Carroll, 2003; Jeong et al., 2006; Kopp et al., 2000; Roeske et al., 2018; Salomone et al., 2013; Signor et al., 2016; Williams et al., 2008).

Identifying potential targets of miRNAs affecting pigmentation

Having identified a collection of miRNAs that affect pigmentation patterning, we next sought to investigate the mechanisms by which these miRNAs might be affecting this phenotype. In canonical miRNA-mediated post-transcriptional repression, miRNAs direct the RNA-induced silencing complex (RISC) to messenger RNAs by binding to short 6-8 base sequences in the 3' UTRs of target genes complementary to a region on the miRNA known as the "seed" (Bartel, 2018). Because the "seed" region of individual miRNAs can be identified by their position in the primary miRNA transcript, potential candidate targets of miRNAs can be identified bioinformatically by scanning the genome for seed matches in the 3' UTRs of protein-coding genes (Enright et al., 2003; Lewis et al., 2003). However, in order to affect the expression of any gene through this canonical mechanism, the miRNA and target mRNA must be co-expressed in the same tissue and developmental stage and there must also be sufficient RISC components present, meaning that a large portion of predicted seed matches in 3' UTRs will not represent biologically relevant direct miRNA-mRNA regulatory connections (Betel et al., 2010). Therefore, we expect only a small proportion of genome-wide predicted seed

sites to be relevant to pigmentation. Unfortunately, detailed expression patterns are unavailable for most *D. melanogaster* miRNAs, so we cannot narrow down potential target genes by filtering for co-expression with miRNAs of interest.

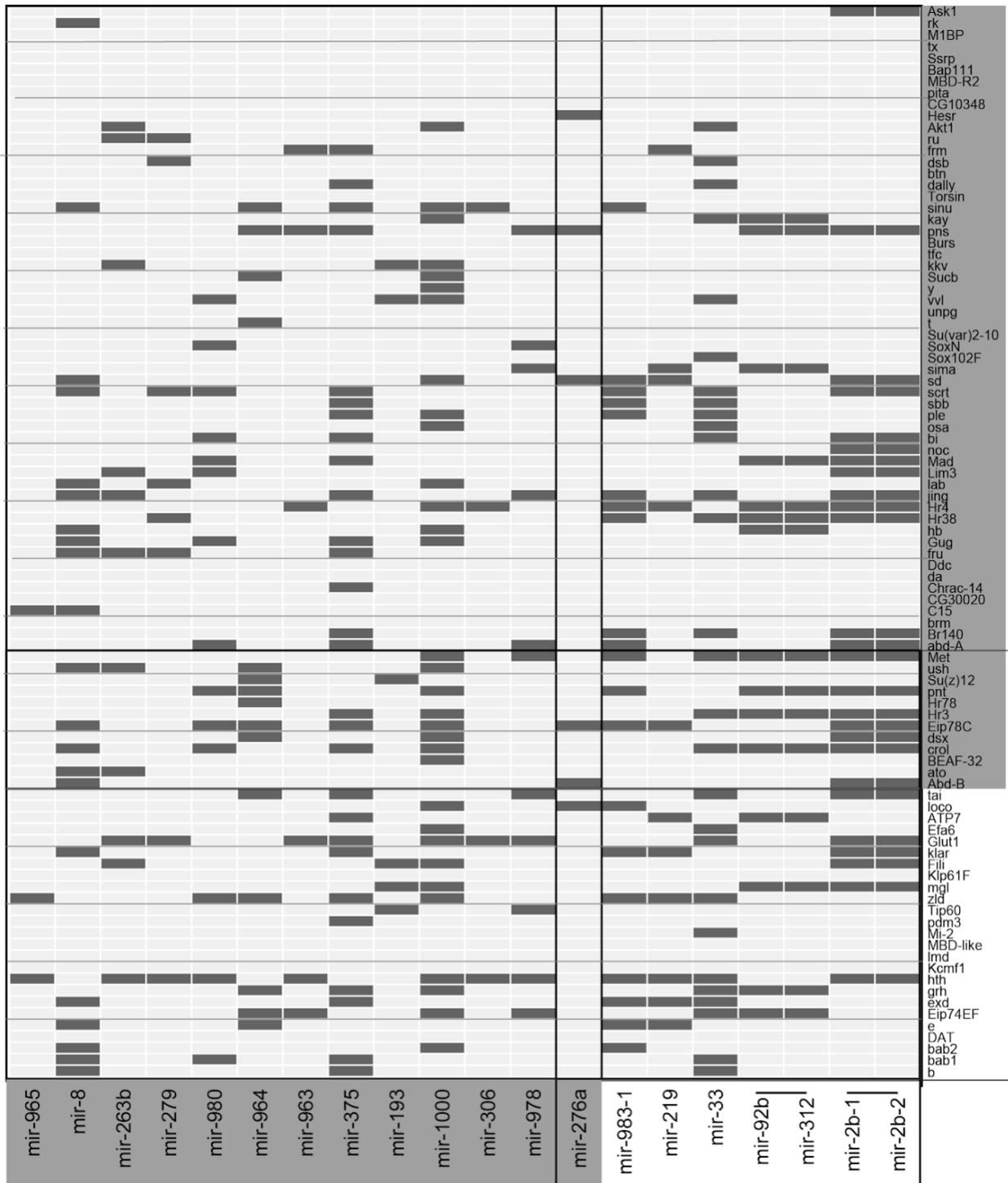


Figure 4- 5. Computationally-predicted seed matches between known pigmentation genes and miRNAs that showed reciprocal phenotypes when overexpressed versus competitively inhibited.

Columns represent miRNAs represented in bold in Table 4-2. miR-2a-1 and miR-2a-2 as well as miR-96b and miR-312 target sites with identical seed sequences according to the TargetScan 7.2 database, and so their predicted targets are identical. Rows represent individual genes from Supplemental Table S4-4, which are known to affect the development of pigmentation in adult cuticle. Light gray boxes surround the names of miRNAs that are sufficient to darken pigmentation in one or more body segments as well as genes that promote the development of dark pigments by the parameters described in the main text. Black-outlined boxes with no fill surround the names of miRNAs that are sufficient to lighten pigmentation in one or more body segments as well as genes that promote the production of light pigments or repress the production of light pigments according to the parameters described in the main text. Several pigmentation genes and miR-276a are contained within both categories, meaning they promote opposite phenotypes in a context-dependent manner. Dark gray rectangles indicate the presence of one or more seed matches to the indicated miRNA within the 3' UTR of the most abundant transcript isoform of the indicated pigmentation gene.

In order to narrow down the predicted miRNA-mRNA pairs and enrich for targets that are likely to affect pigmentation, we identified target predictions for mRNAs from genes that have been shown to affect pigmentation. We generated a list of pigmentation genes by searching gene ontology annotations via flybase.org and manually annotating to include only genes that have been experimentally demonstrated to affect pigmentation, usually through RNAi or loss-of-function mutations (details in Materials and Methods) (Thurmond et al., 2019). We also included genes identified through large-scale screens or GWAS identifying genes underlying pigmentation phenotypes in *D. melanogaster*, so long as the authors provided experimental evidence of the gene's effect on pigmentation (Dembeck et al., 2015a; Kalay et al., 2016; Rogers et al., 2014). We compiled a list of 93 genes that have been experimentally shown to affect pigmentation and annotated their role in pigmentation as “darkens” if the gene has been experimentally demonstrated as either necessary or sufficient for the development of dark pigmentation, and “lightens” if experimental evidence shows that it is necessary to prevent the development of dark pigments or sufficient to lighten pigmentation where misexpressed (Supplemental Table S4-4). If a gene has different effects on pigmentation depending on sex or body segment, we listed its pigmentation role as “context-dependent”. Because the majority of phenotypes we observed in the miRNA overexpression and competitive inhibition screens were located on the female A6 segment, for each gene we listed the effect of the gene's loss-of-function on this segment's phenotype, if known. We also note that for many of the genes in this list, the mechanisms by which they affect pigmentation remain unknown, and it is possible that

the pigmentation effects reported could be caused by disruption to more general developmental processes that alter morphology.

Using the TargetScan 7.2 database of all miRNA 7mer and 8mer seed matches in 3' UTRs of the most-abundant transcript for each gene in the *D. melanogaster* genome (Agarwal et al., 2018), we filtered this dataset to only include the genes in our pigmentation gene list, and further filtered to include only predictions from a set of miRNAs that showed opposite effects on pigmentation when overexpressed versus competitively inhibited in the screens described above (Figure 4-5). Since miRNAs canonically repress their targets, predicted miRNA-target pairs where the miRNA overexpression phenotype (darkening or lightening) is the opposite of the pigmentation gene's role (darkening or lightening) indicate promising candidates for biologically-relevant interaction within the genetic network regulating the development of pigmentation. While we did not see a striking pattern of miRNA-target predictions suggesting that genes which promote dark pigmentation are more likely to contain seed matches for miRNAs which lighten pigmentation, or vice versa, this was not necessarily expected. Presence of a seed site in a gene's 3' UTR is not necessarily sufficient to cause miRNA regulation of that gene for several reasons. Pigmentation develops in the epidermis across several different body segments during a period starting approximately 60 hours after pupation and continuing for several hours after eclosion, meaning there is a wide range of spatial and temporal expression profiles that would be compatible with a gene regulating pigmentation (Bainbridge and Bownes, 1981; Wittkopp et al., 2002). The miRNA and the gene with a predicted seed match may not be co-expressed along with the RISC, meaning the interaction between them would be unlikely to affect pigmentation. Furthermore, even if the expression patterns of the miRNA and predicted target allow for interaction, factors such as the position of the seed within the 3' UTR and the secondary structure of the 3' UTR may leave the seed site less accessible for miRNA binding (Agarwal et al., 2018). These caveats lead to false-positive predictions of miRNA regulation of potential target genes. On the other hand, while more perfect seed matches in 3' UTRs appear to cause the strongest repression by miRNAs, imperfect seed matches and pairing with bases in the mature

miRNA outside the seed region also lead to repression of target genes in *C. elegans* (Broughton et al., 2016). This suggests that there are likely many biologically meaningful miRNA-target interactions occurring that we cannot currently identify by searching for seed matches, leading to an unknown false negative rate.

While the interpretation of predicted miRNA-target binding is difficult, predicted seed pairing may still be useful in generating hypotheses for more direct testing. We note several intriguing miRNA-target predictions in our data that merit further investigation. For instance, miR-276a only had predicted seed sites in 6/93 genes, of which 3 genes (*Abd-B*, *loco*, *sd*) have loss-of-function phenotypes consistent with the miR-276a overexpression phenotype. RNAi against *loco* caused increased pigmentation in female A6 (Dembeck et al., 2015b), while miR-276a overexpression causes increased pigmentation in female A6 (Figure 4-4G,G''). Knockdown of *sd* caused reduced pigmentation in male A2-A6 (Kalay et al., 2016), which we also observed in flies with miR-276a overexpressed (Figure 4-4G'''). *Abd-B* is a most intriguing candidate target of miR-276a, since this gene is a pivotal regulator of sex-specific pigmentation in *D. melanogaster* (Kopp et al., 2000), and miR-276a was the only miRNA we found to have opposite effects on males versus females. RNAi to knock down *Abd-B* causes a loss of male specific pigmentation in A5 and A6, much like the phenotype we observed with miR-276a over-expression (Kalay et al., 2016; Rogers et al., 2014). *Abd-B* is hypothesized to regulate sex-specific pigmentation by binding cis-regulatory sequences associated with *bab1* along with sex-specific isoforms of *dsx*, with *dsx-f* and *Abd-B* together activating *bab1* transcription in the posterior female abdomen, but *dsx-m* and *Abd-B* repressing the activation of *bab1*, which represses the formation of dark pigments (Williams et al., 2008). The overexpression phenotype of miR-276a, increased melanization in female A6 and decreased melanization in Male A5-A6, is consistent with expectations for down-regulating *Abd-B* within this sex-specific genetic network, though further experiments will be needed to determine whether miR-276a affects pigmentation through direct regulation of *Abd-B*. In another notable set of predictions, the hormone receptors *Hr4* and *Hr38*, which both show loss of melanization in female A6 when knocked down with RNAi, each are predicted to be regulated by 4 out of 5 possible

seeds associated with miRNAs that cause only lightened pigmentation with overexpression, while their 3' UTRs contained proportionally fewer seed sites for miRNAs sufficient to only darken pigmentation (*Hr4*, 3/12 seeds, *Hr38* 1/12 seeds, Figure 4-5) (Kalay et al., 2016; Rogers et al., 2014).

Finally, we were interested to assess the mechanism by which the only miRNA previously shown to regulate pigmentation, miR-8, affects pigmentation (Kennell et al., 2012). While miR-8 seed sites were present in 21 pigmentation genes from our list, we noticed that this miRNA has seed matches in 4 genes that are crucial for suppressing melanization of the posterior abdomen in female *D. melanogaster*: *ebony (e)*, *black (b)*, *bric a brac 1 (bab1)*, and *bric a brac 2 (bab2)* (Figure 4-5, Figure 4-6A) (Massey and Wittkopp, 2016b). We were particularly interested in determining whether miR-8 affects pigmentation by targeting a single gene or by coordinately regulating several targets in order to promote the production of dark melanins.

Functional tests suggest miR-8 may coordinately regulate multiple genes involved in the production of light yellow cuticle and suppression of dark melanin synthesis.

We tested whether the predicted seed sites in the 3' UTRs of *e*, *b*, *bab1*, and *bab2* are targeted by *miR-8* by inserting sequences from the 3' UTRs of these potential target genes downstream of the stop codon of a lacZ reporter gene and co-transfecting them with a miR-8 expression vector into S2 cells (Figure 4-6C). We found that reporter genes containing *e*, *b*, *bab1*, and *bab2* 3' UTRs were repressed when co-transfected with miR-8 relative to those co-transfected with an empty vector. When we mutated the predicted miR-8 recognition site in the *e* and *bab1* constructs, we found that the reporter genes were no longer repressed in the presence of miR-8 overexpression, suggesting that these 3' UTRs are directly targeted by miR-8. We observed an unexpectedly large increase in LacZ expression upon deleting the miR-8 seed site in the *bab1* 3'-UTR sequence (Figure 4-6C). We propose several possible reasons for this. First, the sequence cloned into the LacZ expression vector contained 683bp of *bab1* 3'-UTR sequence including the predicted miR-8 seed site, while the full-length *bab1* 3' UTR is

1958bp in length, meaning there could be sequences outside the region we assayed that affect the structure or function of this 3' UTR. Follow-up experiments using the full-length 3' UTR will be necessary to determine whether this is the case. Alternatively, it is possible that ablating the miR-8 seed site in the 3'-UTR sequence we tested may have eliminated miR-8 regulation by endogenous miR-8 in the S2 cells, though we would expect to see the same large increase in expression from the e 3'-UTR construct if both seed sites cause a similar magnitude of repression in the presence of endogenous miR-8 in S2 cells. miR-8 is one of the 10 most abundantly expressed miRNAs in S2 cells (Wessels et al., 2019), suggesting that endogenous miR-8 may be negatively regulating the expression of reporter constructs containing miR-8 seed sites. It is also possible, perhaps likely, that miR-8 overexpression in transfected cells may be altering the expression of endogenous genes, which could affect the expression of other miRNAs with binding sites in the *bab1* 3'-UTR sequences used. In fact, recent data using Photoactivatable Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP) to identify in-vivo miRNA-mRNA interactions in S2 cells identified 52 genes whose transcripts physically associated with miR-8 (Wessels et al., 2019). Further information about the transcriptome-wide effects of over-expressing miR-8 in these cells will be needed to fully explore this possibility. Repeating this experiment in cell lines which lack miR-8 expression, either endogenously or through targeted miR-8 deletion, could also provide clarification.

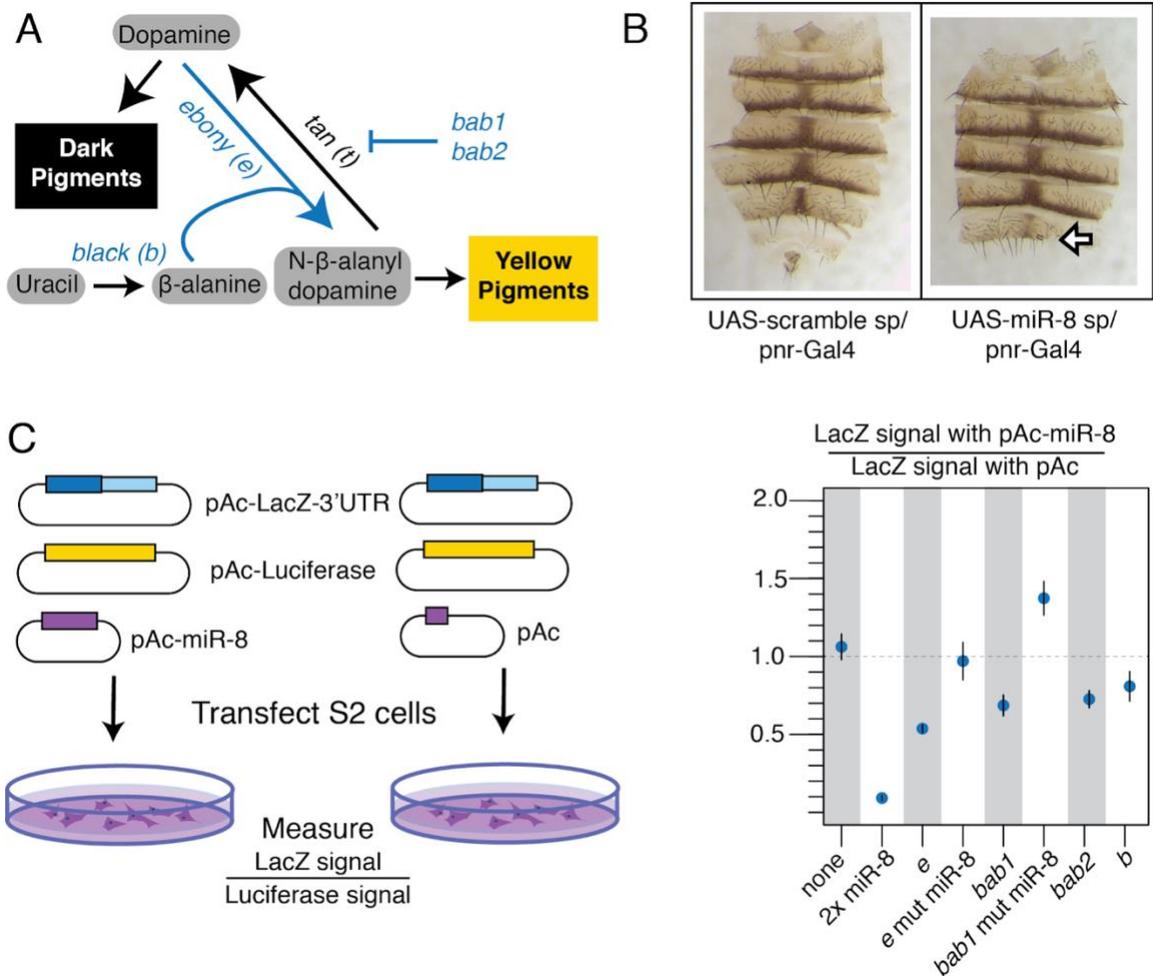


Figure 4-6. miR-8 coordinately regulates *ebony*, *black*, *bric a brac 1* and *bric a brac 2* to promote the production of dark pigments in female A6. **A:** Schematic of a portion of the *D. melanogaster* pigmentation biosynthetic pathway depicting the genetic control of a switch between the production of dark pigments and light, yellow pigments. Metabolites are depicted in gray rounded rectangles. Genes with miR-8 seed sites in their 3' UTRs are shown in blue. Arrowheads represent positive regulation while nail heads represent negative regulation. **B:** Dissected dorsal abdominal cuticle from female flies expressing miRNA sponge transgenes along the dorsal midline. Left: a negative control sponge with scrambled sequence in place of miRNA seed sites. Right: sponge containing miR-8 seed sites. White arrow indicates reduction of pigmentation with competitive inhibition of miR-8. **C:** (Left) Schematic of cell culture assay measuring the effect of miR-8 on the protein production from LacZ reporter genes cloned with potential target gene 3' UTRs. (Right) Relative LacZ signal measured from cells co-transfected with pAc-miR-8 over cells co-transfected with empty pAc vector. A ratio of 1 indicates no difference in LacZ signal in the presence vs absence of pAc-miR-8, while ratios less than or greater than one indicate reduced or increased LacZ signal relative to cells transfected with empty vector in place of pAc-miR-8, respectively. Error bars show 95% confidence intervals. All transfections performed in triplicate. The X-axis labels describe the 3' UTR sequences cloned into each reporter construct downstream of the LacZ stop codon but upstream of the polyadenylation signal. "None" - pAc-LacZ reporter plasmid with no added 3' UTR sequence. "2x miR-8" - a positive control: pAc-LacZ with a 3' UTR composed of two perfect complements to the miR-8 mature sequence. "e" - pAc-LacZ with full wild type *ebony* 3' UTR. "e mut miR-8" - same as "e" construct, but with miR-8 seed site mutated. "bab1" - pAc-LacZ with 683bp of wild type *bab1* 3' UTR containing endogenous miR-8 seed site. "bab1 mut miR-8" - same as "bab1" construct, but with miR-8 seed site mutated. "bab2" - pAc-LacZ with 529bp of wild type *bab2* 3' UTR containing endogenous miR-8 seed site. "b" - pAc-LacZ with full wild type *black* 3' UTR.

Taken together, our results suggest that miR-8 may promote the development of dark pigmentation in the posterior abdomen of female flies by coordinately repressing multiple genes that are each individually necessary to promote the production of yellow pigments rather than dark melanins. In part because miRNAs generally cause mild repression of their individual target genes, rarely exceeding 20% reduction (Guo et al., 2010; Stadler et al., 2012), single miRNAs have long been thought to have little effect on phenotypes (Miska et al., 2007). In contrast to this assertion, changes in expression of a single gene by as little as 29% over a brief period in pupal development contribute to differences in pigmentation between *Drosophila* species to an extent that can be visually distinguished in genetic mapping studies (Cooley et al., 2012). As more studies demonstrate the phenotypic effects of disrupting individual miRNAs, it has become clear that, despite their typically modest direct effects on single targets, many miRNAs are necessary for the development of normal phenotypes in metazoans (Bartel, 2018; Chen et al., 2014; Fulga et al., 2015; Picao-Osorio et al., 2017). One mechanism by which miRNAs might exert their effects on phenotypes is by coordinately regulating multiple target genes with similar functions. In fact, miR-iab-4 and miR-iab-8, which are encoded in a single transcript within the *D. melanogaster* Hox locus, have been shown to coordinately regulate the Hox genes *Abd-A* and *Ubx* as well as the Hox cofactors *exd* and *hth* in the larval ventral nerve cord in a manner that is essential for proper segment patterning, fertility, and mating behavior (Garaulet and Lai, 2015; Garaulet et al., 2014). Our results support a model of combinatorial control, where multiple miRNAs are necessary for wild type pigmentation development, and single miRNAs (such as miR-8) may coordinately regulate multiple functionally related genes.

Conclusion

We propose that the development of *Drosophila* pigmentation is well-suited as a potential model system for the study of miRNA-mediated regulation of gene expression and phenotypic development. Further studies refining the roles of individual miRNAs in the development of pigmentation patterning may be informative to our understanding of the roles of miRNAs, individually and in concert with other miRNAs, within a genetic

network. In addition to the genetic tractability of this model system, *Drosophila* pigmentation displays a wide variety of phenotypic diversity across a wide range of evolutionary time scales. Comparative studies of pigmentation differences within and between species could provide new insights into the evolution of miRNA-mRNA connections within genetic networks.

Materials and Methods

Fly strains and crosses

UAS-miRNA lines were obtained from the FlyORF Zurich ORFeome Project miRNA collection (www.flyORF.ch, Schertel et al, Genetics, 2012). For the overexpression screen, 2 virgin female *pnr*-Gal4/TM6B flies per vial were crossed to male UAS-miRNA flies for each miRNA tested (crosses were set for 172 total UAS-miRNA lines in 8 batches conducted across ~10 weeks). miRNA overexpression crosses were reared in a 23C incubator on a 12hr light/dark cycle. For the competitive inhibition screen, 2 virgin females from the same *pnr*-Gal4 line were crossed to each UAS-miRNA sponge line (obtained from the BDRC - cite Fulga et al sponge paper). Stock information for all UAS-miRNA and UAS-miRNA sponge lines is listed in Supplemental Table S4-3. Because we anticipated potentially smaller effects from competitive inhibition than from overexpression, we set competitive inhibition crosses at two temperature extremes (see main text). We reasoned that the more darkly pigmented flies reared at 18C would allow greater visibility of lightening effects from miRNA inhibition, while lighter flies reared at 28C would allow greater visibility of darkening effects. All flies were aged to 3-5 days after eclosion before phenotyping. The flies' pigmentation outside the *pnr* expression domain was used as an internal control to compare to the effects of miRNA overexpression in the dorsal midline. UAS-*ebony* (Wittkopp et al., 2002) and UAS-*ebony* RNAi (RRID:BDSC 28612) were crossed to *pnr*-

Gal4 as positive controls for lightened and darkened midline pigmentation, respectively. Each fly was observed individually by A. L., at which time the eclosion date, collection date, genotype, sex, and phenotypes were recorded in a spreadsheet for each screen (Supplemental Table S4-1 and Supplemental Table S4-2). Flies with balancer phenotypes were documented with the identity of the balancer. Flies inheriting both *pnr*-Gal4 and a UAS- transgene were documented categorically as unaffected, lightened, darkened, or both lightened and darkened along with a short description of the phenotype and segment(s) affected, as well as any conspicuous developmental defects. Additional crosses were set to compare *pnr*-Gal4/UAS-miRNA flies to a genetically similar negative control by first crossing the *pnr*-Gal4 to a double balancer line (genotype *w⁻;Sco/CyO;MKRS/TM6B*, courtesy of Scott Pletcher) to obtain *pnr*-Gal4/MKRS flies, since, unlike TM6B, MKRS does not include any alleles that visibly affect pigmentation. We crossed *pnr*-Gal4/MKRS to UAS-miR-8, UAS-miR-92b, UAS-miR-33, and UAS-miR-279, rearing and collecting in the same conditions and manner as the UAS-miRNA screen crosses described above.

Sample preparation and imaging

Flies displaying lightened or darkened pigmentation phenotypes were placed in 10% glycerol in ethanol solution for storage and labeled with the associated cross information. In addition, flies from two randomly chosen *pnr*-Gal4/UAS-miRNA crosses for which all individuals were classified as displaying an “unaffected” phenotype (UAS-miR-100 and UAS-miR-1014) were stored in 10% glycerol in ethanol for comparison purposes. A subset of flies with notable phenotypes (including those shown in Figure 4-2) were photographed immediately after collection and before storage, using a Leica MZ6 stereomicroscope equipped with a ring light and Scion (CFW-1308C) camera operated via TWAIN driver in Adobe Photoshop. Dorsal abdominal cuticles were dissected from the offspring of the *pnr*-Gal4/MKRS x UAS-miRNA crosses described in “Fly strains and crosses” for the images in Figure 4-4A-F and Figure 4-6B. Dissection and mounting were performed as described in (John et al., 2016) before imaging. For the images in Figures 4-4G-F, flies preserved from the screen experiments were

partially embedded in 1% agar in white centrifuge tube caps, which were then flooded with 95% ethanol before imaging. All photographs in Figure 4-4 and Figure 4-6 were taken using the same equipment described above.

Analysis of predicted miRNA target genes

We compiled a list of 93 genes with experimentally validated effects on pigmentation by searching literature and annotations from flybase.org (Thurmond et al., 2019). We first compiled the names and phenotypes associated with all genes described in three reviews on *Drosophila* pigmentation (Massey and Wittkopp, 2016b; Wittkopp et al., 2003; Wright, 1987), as well as two large-scale transcription factor RNAi screens for pigmentation regulators (Kalay et al., 2016; Rogers et al., 2014) and one GWAS of within species pigmentation from the *Drosophila* Genome Reference Panel (Dembeck et al., 2015b). We then searched Gene Ontology annotations for “Biological Process” on Flybase.org, filtering for the following terms: “negative regulation of developmental pigmentation”, “positive regulation of developmental pigmentation”, “regulation of cuticle pigmentation”, “regulation of eye pigmentation”, “regulation of female pigmentation”, “regulation of male pigmentation”, “regulation of pigment cell differentiation”, “regulation of adult chitin-containing cuticle pigmentation”. We then checked the references for each gene identified by this search, excluded all that were solely associated with pigmentation in structures other than adult cuticle, and reviewed the evidence supporting any remaining genes that were not described in the previously described sources. We only included genes where the supporting data presented in the referenced studies included descriptions or images of mutant or RNAi phenotypes associated with the gene in question. The results of this search are shown in Supplemental Table S4-4.

For predictions of direct miRNA-mRNA regulatory interactions, we downloaded all predictions for genome wide matches to both “conserved” and “nonconserved” miRNA seed sites in the 3’ UTRs of the most highly expressed transcript of each gene in the *D. melanogaster* genome from TargetScanFly v7.2

(http://www.targetscan.org/fly_72/, (Agarwal et al., 2018)). We then filtered this database to only include only the 3' UTRs of genes listed in Supplemental Table S4-4 and the miRNA seeds associated with miRNAs that produced opposite phenotypes when overexpressed versus competitively inhibited in the screens described in this study (Table 4-2).

Cell culture assays of predicted miR-8 target genes

Reporter genes assays for direct miRNA-mRNA were performed as described in (Blauwkamp et al., 2008) with modifications. For miR-8 expression, the miR-8 gene was cloned into a pAc5.1-V5/His-a expression vector (Invitrogen), while 3' UTRs with an upstream, in-frame stop codon were cloned into a pAc-LacZ vector (Invitrogen) using NotI and XhoI. The full-length 3' UTRs of *e* and *b* were used, while we used a 683bp portion of the *bab1* 3' UTR and a 529bp portion of the *bab2* 3' UTR, each containing the endogenous miR-8 seed site. To ablate miR-8 seed sites, we mutated the 3' UTR sequence of the 7mer, changing all 7 bases from C to A and G to T, or vice versa. We plated 500uL of S2R+ cells in 24-well plates at 1×10^6 cells/mL and allowed them to grow for 24 hours at 25°C before transfection. Cells were transiently transfected using FuGENE 6 transfection reagent (Promega) with a total of 150ng of plasmid DNA (125ng of either pAc-miR-8 or empty vector, 12.5ng of LacZ reporter, and 12.5ng of pAc-Luciferase). Cells were lysed and assayed for Beta-galactosidase and luciferase activity using GalactoStar (Invitrogen) and Tropix LucScreen (Applied Biosystems) approximately 36 hours post-transfection. Beta-galactosidase signal was normalized for transfection efficiency using Luciferase signal. All transfections were performed in triplicate.

References

Agarwal, V., Subtelny, A.O., Thiru, P., Ulitsky, I., and Bartel, D.P. (2018). Predicting microRNA targeting efficacy in *Drosophila*. *Genome Biol.* 19, 152.

- Arif, S., Murat, S., Almudi, I., Nunes, M.D.S., Bortolamiol-Becet, D., McGregor, N.S., Currie, J.M.S., Hughes, H., Ronshaugen, M., Sucena, É., et al. (2013). Evolution of mir-92a underlies natural morphological variation in *Drosophila melanogaster*. *Curr. Biol.* *23*, 523–528.
- Bainbridge, S.P., and Bownes, M. (1981). Staging the metamorphosis of *Drosophila melanogaster*. *J. Embryol. Exp. Morphol.* *66*, 57–80.
- Baker, J.D., and Truman, J.W. (2002). Mutations in the *Drosophila* glycoprotein hormone receptor, rickets, eliminate neuropeptide-induced tanning and selectively block a stereotyped behavioral program. *J. Exp. Biol.* *205*, 2555–2565.
- Bartel, D.P. (2018). Metazoan MicroRNAs. *Cell* *173*, 20–51.
- Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C. (2010). Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* *11*, R90.
- Blauwkamp, T. a., Chang, M.V., and Cadigan, K.M. (2008). Novel TCF-binding sites specify transcriptional repression by Wnt signalling. *EMBO J.* *27*, 1436–1446.
- Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* *118*, 401–415.
- Broughton, J.P., Lovci, M.T., Huang, J.L., Yeo, G.W., and Pasquinelli, A.E. (2016). Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* *64*, 320–333.
- Chen, Y.W., Song, S., Weng, R., Verma, P., Kugler, J.M., Buescher, M., Rouam, S., and Cohen, S. (2014). Systematic study of *Drosophila* MicroRNA functions using a collection of targeted knockout mutations. *Dev. Cell* *31*, 784–800.
- Cooley, A.M., Shefner, L., McLaughlin, W.N., Stewart, E.E., and Wittkopp, P.J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*. *Evol. Dev.* *14*, 317–325.
- Dembeck, L.M., Huang, W., Carbone, M.A., and Mackay, T.F.C. (2015a). Genetic basis of natural variation in body pigmentation in *Drosophila melanogaster*. *Fly* *9*, 75–81.
- Dembeck, L.M., Huang, W., Magwire, M.M., Lawrence, F., Lyman, R.F., and Mackay, T.F.C. (2015b). Genetic Architecture of Abdominal Pigmentation in *Drosophila melanogaster*. *PLoS Genet.* *11*.
- Dewey, E.M., McNabb, S.L., Ewer, J., Kuo, G.R., Takanishi, C.L., Truman, J.W., and Honegger, H.-W. (2004). Identification of the gene encoding bursicon, an insect neuropeptide responsible for cuticle sclerotization and wing spreading. *Curr. Biol.* *14*, 1208–1213.

- Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in *Drosophila*. *Genome Biol.* 5, R1.
- Fulga, T.A., McNeill, E.M., Binari, R., Yelick, J., Blanche, A., Booker, M., Steinkraus, B.R., Schnall-Levin, M., Zhao, Y., DeLuca, T., et al. (2015). A transgenic resource for conditional competitive inhibition of conserved *Drosophila* microRNAs. *Nat. Commun.* 6, 7279.
- Garaulet, D.L., and Lai, E.C. (2015). Hox miRNA regulation within the *Drosophila* Bithorax complex: Patterning behavior. *Mech. Dev.* 138, 151–159.
- Garaulet, D.L., Castellanos, M.C., Bejarano, F., Sanfilippo, P., Tyler, D.M., Allan, D.W., Sánchez-Herrero, E., and Lai, E.C. (2014). Homeotic Function of *Drosophila* Bithorax-Complex miRNAs Mediates Fertility by Restricting Multiple Hox Genes and TALE Cofactors in the CNS. *Dev. Cell* 29, 635–648.
- Garaulet, D.L., Zhang, B., Wei, L., Li, E., and Lai, E.C. (2020). miRNAs and Neural Alternative Polyadenylation Specify the Virgin Behavioral State. *Dev. Cell* 54, 410–423.e4.
- Gibert, J.-M., Peronnet, F., and Schlötterer, C. (2007). Phenotypic plasticity in *Drosophila* pigmentation caused by temperature sensitivity of a chromatin regulator network. *PLoS Genet.* 3, e30.
- Gibert, P., Moreteau, B., and David, J.R. (2004). Phenotypic plasticity of body pigmentation in *Drosophila melanogaster*: genetic repeatability of quantitative parameters in two successive generations. *Heredity* 92, 499–507.
- Gompel, N., and Carroll, S.B. (2003). Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. *Nature* 424, 931–935.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* 466, 835–840.
- Halbeisen, R.E., Galgano, A., Scherrer, T., and Gerber, A.P. (2008). Post-transcriptional gene regulation : From genome-wide studies to principles. *65*, 798–813.
- Jeong, S., Rokas, A., and Carroll, S.B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. *Cell* 125, 1387–1399.
- John, A.V., Sramkoski, L.L., Walker, E.A., Cooley, A.M., and Wittkopp, P.J. (2016). Sensitivity of Allelic Divergence to Genomic Position: Lessons from the *Drosophila tan* Gene. *G3* 6, 2955–2962.
- Kalay, G., Lusk, R., Dome, M., Hens, K., Deplancke, B., and Wittkopp, P.J. (2016). Potential Direct Regulators of the *Drosophila* yellow Gene Identified by Yeast One-Hybrid and RNAi Screens. *G3* 6, 3419–3430.

- Kennell, J. a., Cadigan, K.M., Shakhmantsir, I., and Waldron, E.J. (2012). The microRNA miR-8 is a positive regulator of pigmentation and eclosion in *Drosophila*. *Dev. Dyn.* *241*, 161–168.
- Kopp, A., and Duncan, I. (1997). Control of cell fate and polarity in the adult abdominal segments of *Drosophila* by optomotor-blind. *Development* *124*, 3715–3726.
- Kopp, A., Duncan, I., Godt, D., and Carroll, S.B. (2000). Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* *408*, 553–559.
- Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res.* *47*, D155–D162.
- Lewis, B.P., Shih, I.-H., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. *Cell* *115*, 787–798.
- Massey, J.H., and Wittkopp, P.J. (2016a). The Genetic Basis of Pigmentation Differences Within and Between *Drosophila* Species. *Curr. Top. Dev. Biol.* *119*, 27–61.
- Massey, J.H., and Wittkopp, P.J. (2016b). The Genetic Basis of Pigmentation Differences Within and Between *Drosophila* Species. In *Current Topics in Developmental Biology*, pp. 27–61.
- Miska, E.A., Alvarez-Saavedra, E., Abbott, A.L., Lau, N.C., Hellman, A.B., McGonagle, S.M., Bartel, D.P., Ambros, V.R., and Horvitz, H.R. (2007). Most *Caenorhabditis elegans* microRNAs are individually not essential for development or viability. *PLoS Genet.* *3*, e215.
- Norgate, M., Lee, E., Southon, A., Farlow, A., Batterham, P., Camakaris, J., and Burke, R. (2006). Essential roles in development and pigmentation for the *Drosophila* copper transporter DmATP7. *Mol. Biol. Cell* *17*, 475–484.
- Picao-Osorio, J., Johnston, J., Landgraf, M., Berni, J., and Alonso, C.R. (2015). MicroRNA-encoded behavior in *Drosophila*. *Science* *350*, 815–820.
- Picao-Osorio, J., Lago-Baldaia, I., Patraquim, P., and Alonso, C.R. (2017). Pervasive Behavioural Effects of microRNA Regulation in *Drosophila*. *Genetics* *206*, 1535–1548.
- Rebeiz, M., and Williams, T.M. (2017). Using *Drosophila* pigmentation traits to study the mechanisms of cis-regulatory evolution. *Current Opinion in Insect Science* *19*, 1–7.
- Riedel, F., Vorkel, D., and Eaton, S. (2011). Megalin-dependent yellow endocytosis restricts melanization in the *Drosophila* cuticle. *Development* *138*, 149–158.
- Roeske, M.J., Camino, E.M., Grover, S., Rebeiz, M., and Williams, T.M. (2018). Cis-regulatory evolution integrated the Bric-à-brac transcription factors into a novel fruit fly gene regulatory network. *Elife* *7*.

- Rogers, W. a., Grover, S., Stringer, S.J., Parks, J., Rebeiz, M., and Williams, T.M. (2014). A survey of the trans-regulatory landscape for *Drosophila melanogaster* abdominal pigmentation. *Dev. Biol.* *385*, 417–432.
- Salomone, J.R., Rogers, W. a., Rebeiz, M., and Williams, T.M. (2013). The evolution of *Bab* paralog expression and abdominal pigmentation among *Sophophora* fruit fly species. *Evol. Dev.* *15*, 442–457.
- Schertel, C., Rutishauser, T., Förstemann, K., and Basler, K. (2012). Functional characterization of *Drosophila* microRNAs by a novel in vivo library. *Genetics* *192*, 1543–1552.
- Sekine, Y., Takagahara, S., Hatanaka, R., Watanabe, T., Oguchi, H., Noguchi, T., Naguro, I., Kobayashi, K., Tsunoda, M., Funatsu, T., et al. (2011). p38 MAPKs regulate the expression of genes in the dopamine synthesis pathway through phosphorylation of NR4A nuclear receptors. *J. Cell Sci.* *124*, 3006–3016.
- Shakhmantsir, I., Massad, N.L., and Kennell, J.A. (2014). Regulation of cuticle pigmentation in *Drosophila* by the nutrient sensing insulin and TOR signaling pathways. *Developmental Dynamics* *243*, 393–401.
- Signor, S.A., Liu, Y., Rebeiz, M., Signor, S.A., Liu, Y., Rebeiz, M., and Kopp, A. (2016). Genetic Convergence in the Evolution of Male-Specific Color Patterns in *Drosophila*. *Curr. Biol.* 1–11.
- Stadler, M., Artiles, K., Pak, J., and Fire, A. (2012). Contributions of mRNA abundance, ribosome loading, and post- or peri-translational effects to temporal repression of *C. elegans* heterochronic miRNA targets. *Genome Res.* *22*, 2418–2426.
- Thurmond, J., Goodman, J.L., Strelets, V.B., Attrill, H., Gramates, L.S., Marygold, S.J., Matthews, B.B., Millburn, G., Antonazzo, G., Trovisco, V., et al. (2019). FlyBase 2.0: the next generation. *Nucleic Acids Res.* *47*, D759–D765.
- Verma, P., and Cohen, S.M. (2015). miR-965 controls cell proliferation and migration during tissue morphogenesis in the *Drosophila* abdomen. *Elife* *4*, 1–18.
- Wakabayashi-Ito, N., Doherty, O.M., Moriyama, H., Breakefield, X.O., Gusella, J.F., O'Donnell, J.M., and Ito, N. (2011). Dtorsin, the *Drosophila* ortholog of the early-onset dystonia TOR1A (DYT1), plays a novel role in dopamine metabolism. *PLoS One* *6*, e26183.
- Wessels, H.-H., Lebedeva, S., Hirsekorn, A., Wurmus, R., Akalin, A., Mukherjee, N., and Ohler, U. (2019). Global identification of functional microRNA-mRNA interactions in *Drosophila*. *Nat. Commun.* *10*, 1626.
- Williams, T.M., Selegue, J.E., Werner, T., Gompel, N., Kopp, A., and Carroll, S.B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in *Drosophila*. *Cell* *134*, 610–623.

Wittkopp, P.J., True, J.R., and Carroll, S.B. (2002). Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* 129, 1849–1858.

Wittkopp, P.J., Carroll, S.B., and Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in *Drosophila*. *Trends Genet.* 19, 495–504.

Wright, T.R.F. (1987). The genetics of biogenic Amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*. In *Molecular Genetics of Development*, (Elsevier), pp. 127–222.

Yassin, A., Bastide, H., Chung, H., Veuille, M., David, J.R., and Pool, J.E. (2016). Ancient balancing selection at tan underlies female colour dimorphism in *Drosophila erecta*. *Nat. Commun.* 7, 10400.

Chapter 5

Conclusions and Future Directions

The systems by which genes and genomes encode phenotypes, despite being studied for decades, still remain poorly understood in many ways. Similarly, our understanding of the genetic mechanisms by which phenotypes evolve is limited by many barriers, including gaps in our understanding of genotype-phenotype relationships. While the scientific community's advances in the fields of developmental genetics and evolutionary biology are incrementally accumulating, it has become clear that increased insight into the general principles, exceptions, caveats, mechanisms, and conflicts in either field informs our understanding of the other. In short, investigating how genes and phenotypes evolve informs our understanding of how genes function and affect phenotypes, while our knowledge of how genes encode phenotypes informs the hypotheses we make while investigating the root causes of phenotypic differences.

The studies described in this dissertation aim to approach both sides of this interconnected, multidisciplinary field of inquiry wherein we seek to understand evolution, development, and the evolution of development. In the experiments I describe in the preceding chapters, I sought to improve the methods by which we can directly query the phenotypic effects of natural genetic variants by inducing precise allele replacements (Chapter 2), expand our understanding of the genetic basis of an evolved difference between two recently diverged species (Chapter 3), and broaden our understanding of the genetic architecture underlying the development of a model trait to include a class of molecules that had previously been largely unexplored in the context of this trait (Chapter 4). In this chapter, I describe the conclusions drawn from the

research presented in this body of work, as well as potential future directions for addressing the questions that arise from this research.

Improving the efficiency of precise insertions of DNA sequence at a gene's native locus brings feasibility to previously intractable strategies for the study of genetics and evolution.

In Chapter 2, I describe a 2-stage allele replacement method in *Drosophila melanogaster*, with an emphasis on preventing unwanted secondary mutations and only using materials that are not limited to specific transgenic lines. This method was also developed to overcome the widely-reported problem of low-frequency of CRISPR-induced incorporation of donor DNA at the locus of interest (Liu et al., 2017), which often necessitates laborious, costly screening through large numbers of potential mutants to identify rare successes. While I was ultimately able to execute an allele replacement without unwanted secondary mutations using this method, there is certainly room for improvement. While I initially aimed to replace the intermediate reporter construct with a series of three co-injected alleles, I ultimately only recovered mutants containing one of these alleles. I propose that this was a result of the overall low efficiency of homology-directed repair we observed, and that multiple allele replacements could be induced within a single round of injections if one could identify the experimental parameters necessary to increase the rate of donor DNA incorporation at the locus being edited.

Several approaches have been successful in increasing the efficiency of precise DNA sequence changes via homology-directed repair. Some particularly efficient methods described since the publication of the research in Chapter 2 involve tethering the donor DNA repair template to the CRISPR machinery, for instance by covalently attaching a multi-copy single-stranded DNA donor to a Retron scaffold on the sgRNA in Cas9 Retron precise Parallel Editing via homology (CRISPEY) ((Sharon et al., 2018), or by tethering the single-stranded DNA donor to a fusion protein made up of Cas9 and HUH endonuclease, which covalently binds ssDNA (Aird et al., 2018). While these

tethering methods achieve great efficiency increases in recovering precise integration of sequences from single-stranded DNA donors, their applications may be limited. For instance, the CRISPEY method was developed in yeast, and involved the use of yeast strains with genomically integrated transgenes expressing reverse transcriptase and Cas9, meaning significant changes would be needed to adapt this method for use in non-transgenic animal models, for instance (Sharon et al., 2018). In addition, tethering methods make use of single stranded DNA donors of approximately 100 to 200nt in length including homology regions, and it is unclear at present whether these methods will be adaptable to inserting longer sequences. However, improvement of homology-directed repair efficiency is an area of active, rapid research, and many other approaches, such as drug-based or genetic manipulations of endogenous DNA repair machinery (Ray and Raghavan, 2020), show promise for applicability across a wider range of model systems.

Investigating the genetic causes of evolved differences in pigmentation: an opportunity to test hypotheses about pleiotropy and modularity.

In Chapter 3, I directly assessed the role of *ebony* in the evolved light body color of *Drosophila novamexicana* relative to its sister species *D. americana* and all other known species in the Virilis group. Studies into this pigmentation difference identified *ebony* as a candidate gene initially in 2003, as it resides in a quantitative trait locus (QTL) strongly associated with the trait of interest, but chromosomal rearrangements prevented any direct assessment of the effects of *ebony* without linked sequences confounding interpretation (Cooley et al., 2012; Wittkopp et al., 2003, 2009). By using CRISPR/Cas9 genome editing to generate *ebony* null mutants in both *D. novamexicana* and *D. americana* and comparing the phenotypes of reciprocal F1 hybrids differing only in the species of origin of their intact *ebony* allele, I provided the first direct evidence that divergence at *ebony* is responsible for a large portion of the pigmentation difference between these species.

Evidence collected over years of study suggests that the effect attributed to the *ebony*-containing QTL acts in *cis*, or on the same chromosome rather than through the action of diffused factors, and not through coding changes in *ebony* (Cooley et al., 2012; Wittkopp et al., 2009). As described in Chapter 1, evolution through changes in non-coding *cis*-regulatory sequence are thought to be evolutionarily favorable in cases where the associated gene is expressed in a modular fashion and has pleiotropic effects on multiple phenotypes (Carroll, 2005; Stern and Orgogozo, 2008). While *ebony* fits these qualifications, as it is required in the nervous system for neurotransmitter recycling and vision (Borycz et al., 2002), affects the production of cuticular hydrocarbons (Massey et al., 2019a), and has demonstrated modular *cis*-regulatory elements driving expression in different tissues (Rebeiz et al., 2009), it is unclear whether changes at *ebony* that affect pigmentation are limited to expression in the developing cuticle or are also causing changes in other *ebony*-dependent phenotypes. To begin to address whether the effects of differences in *ebony* were limited to body color, we assessed the effect of both species' *ebony* alleles on cuticular hydrocarbons and found that, while *ebony* does indeed affect cuticular hydrocarbons in these species, there was no detectable difference in the effects of the two species' *ebony* alleles on this phenotype. While we cannot determine whether the effects of differences in *ebony* function between these alleles was limited to changing pigmentation, the absence of an allele-specific effect on cuticular hydrocarbons at least suggests that evolution at *ebony* in these species did not have widespread, detectable effects on all phenotypes that rely on *ebony*.

Interestingly, in all cases of pigmentation evolution within *Drosophila* where the causative sequence changes have been identified, the mutations are non-coding, *cis*-regulatory changes (Massey and Wittkopp, 2016). While this model system affords us the rare opportunity to investigate the extent of phenotypic changes resulting from evolution of *cis* regulatory sequences that control the expression of genes with known pleiotropic roles in multiple traits (Takahashi, 2013; Wittkopp and Beldade, 2009), studies rarely investigate whether the mutations underlying pigmentation differences affect expression in other tissues associated with other characteristics. I see this as a

missed opportunity to test the hypothesis that *cis*-regulatory changes are favored because of their limited pleiotropy. Furthermore, investigating the effects of known alleles with different effects on pigmentation outside of the tissues and developmental stages related to pigmentation has the potential to possibly aid in the investigation of the selective forces underlying changes in pigmentation. It is possible that pigmentation differences could, in some cases, be tied to changes in cuticular hydrocarbons, vision, behavior, or other characteristics that have been shown to rely on genes such as *yellow*, *ebony*, and *tan* that have repeatedly been found to underlie differences in pigmentation (Borycz et al., 2002; Kennell et al., 2012; Massey et al., 2019a, 2019b; Wittkopp and Beldade, 2009).

More mechanistic studies into the evolution of increased expression are needed.

While the experiments described in Chapter 3 demonstrate the long-hypothesized role of *ebony* in the evolution of light body color in *D. novamexicana* and *D. americana*, we still do not know the specific mutations that cause this change in *ebony* function, nor the mechanism by which these mutations change *ebony* activity. While it has been shown that *D. novamexicana* *ebony* is more highly expressed than *D. americana* *ebony* during the stages when pigmentation develops (Cooley et al., 2012), it is unknown how *D. novamexicana* *ebony* gained this increase in expression. Molecular insight into the increase in *D. novamexicana* *ebony* expression is a particularly attractive prospect, considering that of the few cases where the mutations underlying a change in pigmentation have been identified and characterized, most cause a decrease in the expression of the associated gene (Jeong et al., 2008; Martin and Orgogozo, 2013; Prud'homme et al., 2006). In fact, this trend has been observed beyond *Drosophila*, leading to a general hypothesis that tissue-specific loss of expression through *cis*-regulatory mutation may be a particularly common mode of phenotypic evolution (Chan et al., 2010; Gompel and Prud'homme, 2009; Martin and Orgogozo, 2013; Prud'homme et al., 2007; Rus et al., 2005; Stern and Frankel, 2013). Interestingly, another case of

derived light body pigmentation in *Drosophila*, this time within-species variation in *D. melanogaster*, was found to result from an increase in *ebony* expression, but this increase in expression was due to loss-of-function in a silencer element (Johnson et al., 2015). The overall trend in the available case studies suggest that evolution of phenotypes through *cis*-regulatory changes may occur largely through loss-of-function mutations in regulatory sequence, even in cases where expression is increased. It would be interesting to determine whether the increase in *ebony* expression in *D. novamexicana*, too, is caused by a loss of silencer function, or by other means. Improvements in genome editing methods to swap sequences at native loci in non-model *Drosophilids* would greatly facilitate mechanistic studies of this locus, allowing the exchange of candidate regulatory sequences between *D. novamexicana* and *D. americana* *ebony* alleles to identify the causative mutations.

Gene expression is more than just transcription: Incorporating post-transcriptional regulation into the study of evolution and development in pursuit of a more holistic understanding.

While Chapters 2 and 3 of this dissertation describe specific, targeted efforts to identify the genetic basis of phenotypic differences and to develop genetic tools to facilitate this process, the motivations behind the research described in Chapter 4 were more broadly exploratory. Noting a distinct dearth of available information on the role of post-transcriptional regulation in the evolution of genetic regulation (Chen and Rajewsky, 2007), I set out to determine whether microRNAs (miRNAs), a class of small, non-coding RNAs that canonically repress messenger RNAs before translation, play a role in the development of *D. melanogaster* pigmentation. I found that miRNAs do, in fact, regulate the development of this trait, presenting evidence for 22 miRNAs that appear to be necessary for the development of wild type pigmentation phenotypes in *D. melanogaster*. Furthermore, I provide evidence that miR-8, the one miRNA previously shown to affect pigmentation development in this species, may exert its effects on pigmentation by coordinately regulating several protein coding genes with long-

established roles in promoting the production of light yellow pigments over dark melanins. It is my hope that the results presented in Chapter 4, in addition to providing insight into the roles of miRNAs in the development of *D. melanogaster* pigmentation, might serve as evidence that this model system is a potentially valuable source of insight into the roles of miRNAs in development and evolution more generally. There are many areas of inquiry that I would like to see explored as a follow-up to my findings on this subject.

Understanding the role of miRNAs within the D. melanogaster pigmentation developmental network: next steps

The overexpression and competitive inhibition screens described in Chapter 4 revealed that dozens of miRNAs are sufficient to cause either darkened or lightened pigmentation phenotypes in the dorsal cuticle of the abdomen or thorax, and 22 of the 41 miRNAs competitively inhibited caused changes in pigmentation, suggesting that these miRNAs are necessary for the development of normal pigmentation patterns. In order to illuminate the roles of these miRNAs within the genetic network underlying pigmentation development, further investigation into the genes targeted by these miRNAs is necessary. In the case of miR-8, for instance, the reporter assays I performed in cell culture provide evidence that the 3' untranslated regions (UTRs) of *ebony*, *black*, *bric a brac 1* and *bric a brac 2* are responsive to strong co-expression of miR-8. However, there is currently no direct evidence that any of these genes are regulated by miR-8 in vivo during the stages when pigmentation develops. Several avenues of inquiry could clarify this. First, immunofluorescence against putative targets of miR-8 could be performed in the presence of miR-8 overexpression, competitive inhibition of miR-8 via sponge transgene, wild type miR-8 expression, and a miR-8 mutant background. Functional targets of miR-8 should show reduced protein abundance with miR-8 overexpression (unless endogenous miR-8 expression is sufficient to fully repress the target), while miR-8 inhibition or loss of function should cause increased protein abundance of functional targets. While these experiments

would demonstrate that miR-8 affects these target genes, however, it would not be able to distinguish between direct and indirect regulation since the protein abundance of the potential target gene may be affected by other changes in the genetic environment in response to miR-8 presence or absence. Ideally, precise mutation of the predicted miR-8 binding sites in the 3' UTRs of these genes could be induced using CRISPR-Cas9, after which manipulations of miR-8 presence could be performed as described above, with the expectation that ablation of the seed site in a direct target gene's 3' UTR will cause the protein abundance of the target gene to no longer respond to these manipulations.

Alternatively, or in addition, direct assays of miRNA-target interaction may be employed, such as Cross-Linking ImmunoPrecipitation (CLIP) or Chimera PCR (ChimP) (Broughton and Pasquinelli, 2018; Wessels et al., 2019). These methods have the benefit of providing an opportunity to investigate all miRNA-target interactions present in the assayed tissue sample rather than focusing on a single, specific miRNA-target pair. However, care must be taken in planning these experiments and interpreting the results. Previous research has shown that genes can exert their effects on pigmentation during short, transient time frames, making it easy to miss biologically relevant information depending on the developmental staging of specimens (Cooley et al., 2012). Furthermore, the evidence of a physical interaction between a miRNA and 3' UTR of a mRNA is not sufficient to prove the miRNA is repressing the translation of the mRNA to which it is bound. For instance, miRNA-target pairs where binding occurred primarily via sequence outside the canonical "seed" region of the miRNA have been identified in CLIP data but shown to have no detectable effect on transcript abundance of the targeted genes, while other classes of miRNA-target interactions generally reduced mRNA abundance (Broughton et al., 2016). Therefore, the caveats and limitations of each method must be carefully considered when determining a strategy to identify biologically meaningful miRNA-target interactions.

Models of genetic architecture in metazoans lacking post-transcriptional information are incomplete.

Beyond investigating the direct relationships between miRNAs and target genes that affect the development of pigmentation, the timing and patterning of miRNA expression within the developing cuticle is also an interesting area of inquiry. Importantly, well-characterized expression profiles for miRNAs will immensely aid in narrowing down potential target genes through which they may affect pigmentation by cross-referencing their expression patterns against those of predicted targets (Luo et al., 2019). In addition, spatial and temporal expression patterns of miRNAs that affect pigmentation might be informative as to how they regulate the development of pigmentation. Genes such as *yellow*, *tan*, and *ebony* have expression patterns that prefigure the patterning of adult pigmentation patterns, with strong *ebony* expression in regions that will become yellow cuticle, and *yellow* and *tan* expression in areas that will be dark/black in adult cuticle (Ordway et al., 2014; Rebeiz et al., 2009; Wittkopp et al., 2002a). If any miRNAs with demonstrated effects on pigmentation have expression patterns similar to known pigmentation genes or reminiscent of adult pigment patterns, this would raise interesting questions. Are the miRNAs expressed in the developing cuticle regulated by the same transcription factors as other pigmentation genes that are expressed in similar patterns? Do the expression patterns of miRNAs that affect pigmentation change across strains or species of flies with differing pigmentation patterns as is often seen with protein-coding genes like *ebony*, *tan*, *yellow*, and *bric a brac* (Ordway et al., 2014; Salomone et al., 2013; Wittkopp et al., 2002b)? A significant barrier to investigating the expression patterns of miRNA expression, however, is the technical difficulty of visualizing the expression of small, non-coding RNAs which, unlike proteins cannot be detected with antibodies, and unlike mRNAs cannot be easily targeted with nucleic acid probes using traditional methods due to their short length (Urbanek et al., 2015). However, a new, highly sensitive in-situ hybridization method called hybridization chain reaction has recently been shown to detect miRNAs in

zebrafish and mouse, suggesting that this method may be applicable in *Drosophila* as well (Acharya, 2016; Zhuang et al., 2020).

Characterizing the roles of miRNAs within the genetic network underlying *D. melanogaster* pigmentation may offer rare insights into the regulatory logic coordinating the expression of miRNAs, their target genes, and the transcription factors that regulate the expression of both of these classes of genes. A few case studies have identified interesting regulatory relationships involving miRNAs, such as transcription factors regulating the miRNAs that target their own 3' UTRs, suggesting a feedback mechanism (Finnegan and Pasquinelli, 2013; Martinez et al., 2008). Coordinated expression of two miRNAs with distinct seed sequences encoded on complementary strands of the same locus has also been described in two separate cases, with miRNAs expressed from both strands affecting the same biological processes (Garaulet et al., 2014; Scott et al., 2012). Intriguingly, I found that the paralogs *bric a brac 1* and *bric a brac 2*, which both repress the production of black pigment in the posterior abdomen of female *D. melanogaster* in a sexually dimorphic regulatory circuit, each contain seed matches to miR-8 that are responsive to miR-8 in cell culture assays, despite the fact that the 3' UTRs did not have overlapping predicted binding sites for other miRNAs in my analysis (see Chapter 4, Figure 4-5 and Figure 4-6) (Couderc et al., 2002; Kopp et al., 2000). It would be interesting to investigate the evolution of 3' UTR sequences and miRNA regulation of paralogs over evolutionary time. While *bric a brac 1* and *bric a brac 2* are co-expressed in most *Drosophila* lineages, divergent expression of the two paralogs has been identified in at least one lineage (Salomone et al., 2013). These paralogs could provide a system to compare the evolutionary rates of transcriptional and post-transcriptional regulation over a broad range of evolutionary distances. Consideration of both transcriptional regulation and post-transcriptional regulation of genes can provide a more complete understanding of the regulation of gene expression. It is worth noting that common approaches to characterizing the expression patterns of genes using reporter genes, such as enhancer traps or transgenes composed of enhancer or promoter sequences cloned upstream of reporters, separate the gene from any regulatory information contained within the gene's native 3' UTR. This could plausibly

cause reporter expression to occur outside of the regions where the native protein would be expressed if it is usually repressed by miRNAs through sequences in its 3' UTR. Interestingly, ectopic expression has been identified in many enhancer trap lines (Casas-Tintó et al., 2017). I propose that the *D. melanogaster* pigmentation model system provides an excellent opportunity to investigate the holistic regulatory architecture and evolution of gene expression including the combined actions of pre- and post-transcriptional mechanisms.

References

- Acharya, A. (2016). Multiplexed Analysis of Diverse RNA Classes via Hybridization Chain Reaction. phd. California Institute of Technology.
- Aird, E.J., Lovendahl, K.N., St Martin, A., Harris, R.S., and Gordon, W.R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun Biol* 1, 54.
- Borycz, J., Borycz, J. a., Loubani, M., and Meinertzhagen, I. a. (2002). tan and ebony genes regulate a novel pathway for transmitter metabolism at fly photoreceptor terminals. *J. Neurosci.* 22, 10549–10557.
- Broughton, J.P., and Pasquinelli, A.E. (2018). Detection of microRNA-Target Interactions by Chimera PCR (ChimP). *Methods Mol. Biol.* 1823, 153–165.
- Broughton, J.P., Lovci, M.T., Huang, J.L., Yeo, G.W., and Pasquinelli, A.E. (2016). Pairing beyond the Seed Supports MicroRNA Targeting Specificity. *Mol. Cell* 64, 320–333.
- Carroll, S.B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3, e245.
- Casas-Tintó, S., Arnés, M., and Ferrús, A. (2017). Drosophila enhancer-Gal4 lines show ectopic expression during development. *Royal Society Open Science* 4, 170039.
- Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr, Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J., et al. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327, 302–305.
- Chen, K., and Rajewsky, N. (2007). The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.* 8, 93–103.

- Cooley, A.M., Shefner, L., McLaughlin, W.N., Stewart, E.E., and Wittkopp, P.J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in *Drosophila americana* and *D. novamexicana*. *Evol. Dev.* 14, 317–325.
- Couderc, J.-L., Godt, D., Zollman, S., Chen, J., Li, M., Tiong, S., Cramton, S.E., Sahut-Barnola, I., and Laski, F. a. (2002). The *bric à brac* locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in *Drosophila*. *Development* 129, 2419–2433.
- Finnegan, E.F., and Pasquinelli, A.E. (2013). MicroRNA biogenesis: regulating the regulators. *Crit. Rev. Biochem. Mol. Biol.* 48, 51–68.
- Garaulet, D.L., Castellanos, M.C., Bejarano, F., Sanfilippo, P., Tyler, D.M., Allan, D.W., Sánchez-Herrero, E., and Lai, E.C. (2014). Homeotic Function of *Drosophila* Bithorax-Complex miRNAs Mediates Fertility by Restricting Multiple Hox Genes and TALE Cofactors in the CNS. *Dev. Cell* 29, 635–648.
- Gompel, N., and Prud'homme, B. (2009). The causes of repeated genetic evolution. *Dev. Biol.* 332, 36–47.
- Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., and Carroll, S.B. (2008). The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* 132, 783–793.
- Johnson, W.C., Ordway, A.J., Watada, M., Pruitt, J.N., Williams, T.M., and Rebeiz, M. (2015). Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between *Drosophila* Species. *PLoS Genet.* 11, e1005279.
- Kennell, J. a., Cadigan, K.M., Shakhmantsir, I., and Waldron, E.J. (2012). The microRNA miR-8 is a positive regulator of pigmentation and eclosion in *Drosophila*. *Dev. Dyn.* 241, 161–168.
- Kopp, A., Duncan, I., Godt, D., and Carroll, S.B. (2000). Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* 408, 553–559.
- Liu, Z., Liang, Y., Ang, E.L., and Zhao, H. (2017). A New Era of Genome Integration—Simply Cut and Paste! *ACS Synth. Biol.* 6, 601–609.
- Luo, J., Pan, C., Xiang, G., and Yin, Y. (2019). A Novel Cluster-Based Computational Method to Identify miRNA Regulatory Modules. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 16, 681–687.
- Martin, A., and Orgogozo, V. (2013). The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* 67, 1235–1250.
- Martinez, N.J., Ow, M.C., Barrasa, M.I., Hammell, M., Sequerra, R., Doucette-Stamm, L., Roth, F.P., Ambros, V.R., and Walhout, A.J.M. (2008). A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes Dev.* 22, 2535–2549.

Massey, J.H., and Wittkopp, P.J. (2016). The Genetic Basis of Pigmentation Differences Within and Between *Drosophila* Species. In *Current Topics in Developmental Biology*, pp. 27–61.

Massey, J.H., Akiyama, N., Bien, T., Dreisewerd, K., Wittkopp, P.J., Yew, J.Y., and Takahashi, A. (2019a). Pleiotropic Effects of ebony and tan on Pigmentation and Cuticular Hydrocarbon Composition in *Drosophila melanogaster*. *Front. Physiol.* *10*, 518.

Massey, J.H., Chung, D., Siwanowicz, I., Stern, D.L., and Wittkopp, P.J. (2019b). The yellow gene influences *Drosophila* male mating success through sex comb melanization. *Elife* *8*, 1–20.

Ordway, A., Hancuch, K.N., Johnson, W., Williams, T.M., and Rebeiz, M. (2014). The expansion of body coloration involves coordinated evolution in cis and trans within the pigmentation regulatory network of *drosophila prostipennis*. *Dev. Biol.* 1–10.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. a., Williams, T.M., Yeh, S.-D., True, J.R., and Carroll, S.B. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature* *440*, 1050–1053.

Prud'homme, B., Gompel, N., and Carroll, S.B. (2007). Emerging principles of regulatory evolution. *Proc. Natl. Acad. Sci. U. S. A.* *104 Suppl*, 8605–8612.

Ray, U., and Raghavan, S.C. (2020). Modulation of DNA double-strand break repair as a strategy to improve precise genome editing. *Oncogene* *39*, 6393–6405.

Rebeiz, M., Pool, J.E., Kassner, V. a., Aquadro, C.F., and Carroll, S.B. (2009). Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population. *Science* *326*, 1663–1667.

Rus, A., Baxter, I., Muthukumar, B., Gustin, J., Lahner, B., Yakubova, E., and Salt, D.E. (2005). Natural Variants of AtHKT1 Enhance Na Accumulation in Two Wild Populations of *Arabidopsis*. *PLoS Genetics preprint*, e210.

Salomone, J.R., Rogers, W. a., Rebeiz, M., and Williams, T.M. (2013). The evolution of Bab paralog expression and abdominal pigmentation among *Sophophora* fruit fly species. *Evol. Dev.* *15*, 442–457.

Scott, H., Howarth, J., Lee, Y.B., Wong, L.-F., Bantounas, I., Phylactou, L., Verkade, P., and Uney, J.B. (2012). MiR-3120 is a mirror microRNA that targets heat shock cognate protein 70 and auxilin messenger RNAs and regulates clathrin vesicle uncoating. *J. Biol. Chem.* *287*, 14726–14733.

Sharon, E., Chen, S.-A.A., Khosla, N.M., Smith, J.D., Pritchard, J.K., and Fraser, H.B. (2018). Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell* *175*, 544–557.e16.

- Stern, D.L., and Frankel, N. (2013). The structure and evolution of cis-regulatory regions: the shavenbaby story. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20130028.
- Stern, D.L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* 62, 2155–2177.
- Takahashi, A. (2013). Pigmentation and behavior: potential association through pleiotropic genes in *Drosophila*. *Genes Genet. Syst.* 88, 165–174.
- Urbanek, M.O., Nawrocka, A.U., and Krzyzosiak, W.J. (2015). Small RNA Detection by in Situ Hybridization Methods. *Int. J. Mol. Sci.* 16, 13259–13286.
- Wessels, H.-H., Lebedeva, S., Hirsekorn, A., Wurmus, R., Akalin, A., Mukherjee, N., and Ohler, U. (2019). Global identification of functional microRNA-mRNA interactions in *Drosophila*. *Nat. Commun.* 10, 1626.
- Wittkopp, P.J., and Beldade, P. (2009). Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy. *Seminars in Cell and Developmental Biology* 20, 65–71.
- Wittkopp, P.J., True, J.R., and Carroll, S.B. (2002a). Reciprocal functions of the *Drosophila* yellow and ebony proteins in the development and evolution of pigment patterns. *Development* 129, 1849–1858.
- Wittkopp, P.J., Vaccaro, K., and Carroll, S.B. (2002b). Evolution of yellow gene regulation and pigmentation in *Drosophila*. *Curr. Biol.* 12, 1547–1556.
- Wittkopp, P.J., Williams, B.L., Selegue, J.E., and Carroll, S.B. (2003). *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1808–1813.
- Wittkopp, P.J., Stewart, E.E., Arnold, L.L., Neidert, A.H., Haerum, B.K., Thompson, E.M., Akhras, S., Smith-Winberry, G., and Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*. *Science* 326, 540–544.
- Zhuang, P., Zhang, H., Welchko, R.M., Thompson, R.C., Xu, S., and Turner, D.L. (2020). Combined microRNA and mRNA detection in mammalian retinas by in situ hybridization chain reaction. *Sci. Rep.* 10, 351.

Appendices

Appendix A
Supplemental Data for Chapter 2

Supplemental Text S2-1: Annotated pGEM-WingGFP-*tan* sequence (7290bp):

Underlined sequences represent primers used for plasmid construction and screening.

Double underlines represent areas where two primers overlap

5' homology arm

5' target PAM

Unique portion of 5' target site for reporter excision (restriction sites: BglII, BsiWI, Acc65I)

Pupal wing enhancer from *D. melanogaster yellow* gene

EGFP with Hsp70 promoter and SV40 polyadenylation signal

Unique portion of 3' target site for reporter excision (restriction sites: NarI, Bsu36I, ClaI)

3' homology arm

3' target PAM

Restriction sites within pGEM backbone for homology arm removal: 5' end – ApaI, SphI; 3' end – SbfI, MluI

ATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTC
ACGACGTTGTA AACGACGCGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGGGCCGACGTCGCA
TGCTCCCGGCCGCCATGGCGGCCGCGGAATTCGATCTCTGGATAAGCGTCAGCCTCGCGTTGCGTGCAGT
CGCAGTCGCAGTCGCCTTCGCTGCTGGCGTCGCAGCGTCGCCATTGCTGTTATTGTTGTTGCTGTGTTTCTAT
ATGCGTTGCCTGCTCCGTTATCGCAGCGACGTCGTCGTCGTCGTCGTCGTCATCAGCTGTCTTCGTGTTGT
TGTTGTGCTTTTTGCTGAGTGTGTTGTTTCGGCACCTCGAATTGCTGCCGCAAACATGTCCAATTCAGTTT
GAATTCTGCCTCAAGCGCTTTAGAACAATCATATAAAAACCTGGCTTGAATTATCTATATACAATTTGTTTACA
CTGAAATTCTCTCTCTCTCGTAACCTTTCTCTCGCTCTGTCTGTGTGTGCGTGCCTGTGTTGAGTGTGG
TAAAGTTGTTAATGAAATTATTAACCTCAGTTGTGTTTTTTTTTATTATTATTATTATTCATCTTGTGAATTTG
TTTTGTTTTGTTGAAAATTTGTTGTAATTCATTTCCAGCGAGTTTTATTCCCACTGAATTATTATATATAT
TTCTGTGCTTTATTTTGTTCCTTTCTATTTACATTTGTGATCGTGGATAAGACGAACCAAGCAAAAA
AAAAAAAAGAAAAGAAAAGAATAATTGAGAGCATATTACTGCTCATCCTGCTCGAATTTAGAGCCAAAT
AGCGCGAAGAGCGATAAACGGTCCACAGCACATAGACCAAACGAAAGCCAGGAAGCCATGTGCACCTCC
CCGGTGGCCACCAAATGTCGTCCAGCATGTCCTCCGGCAAGATCTCGTACGGTACCATCCATACTCATCAAT
TATCTTATAGCTTATCGTTACGTGTACGGCGCGCAATGGCGAAACTGGAAACACTAACTAGTAAATGCGAA

ATGTATCATATGGATTTTCATATTTTTCCACTGCTAATGGAAATGTTTGCCTTGAAGAGATTGGTCGACTATTA
AATGATTATCGCCCGATTACCACATTGAGTGGTTTAAAAATAGCCATAAAATATGCAACTGACGATGGCTTAAG
ATAAATACGTCGCAGAGTCACTCATAAATTTGAAACGCAGCCCGCTGATTTACCTACCCCTCTAAACGATTTCAT
AGTATATGTACGAGTATATCCACTAAGCTTTTTCGAGCACTGATTTTTTCGCTTGCACGAGACAAGTGCACCAC
CGCAATTGCAGGCAAATTATGTCTGAGGTAATGATTCGGTTTCGTGCAAGATTACACAGAAATCAAATTACGA
CAACCTTTATTCAGTAAGCAAACAAAGCCTTTGTTGGCATCTAATTATCCACTTATGGTTGCGATTTCCGGGAG
CTACAATCGGTTTTGGTTTAGTATATCTAGCGAGTTCCTTGGCGACATTTAAAATTTACAAATAAAGTTTCTCTA
TTCAATCGGGACAGTGGAAATTGACTATTTTATTTATTTAATGAACTTATTTTAAATTTGGCTTAAGTTACTAA
GGGGTACTAATAGTTTGGAGCGCAGTGCATGTCATGGGGACATGTGCAATTGTGTGTAAGCGGGAAGTGATC
GCGGCCTTCCGAATTTGGCCATGCCAAATAATCCAGCTCGAAAGGAGGGGACCCGGCGGTTCAGGGCCATG
GACATTGAACTTGAAAAAAAAAAAAACACAAAAATATATAACACAAAAACGGAAAATGCTGTGTACCCTTATGT
TAGAGAAGTTGAGCAACGGGTTTTTCGTTTTGCAGTCCAGTGGATTTCCAAATTAGTGTAGGAGGGGGGAG
GGGAGGGAGGGAGATAATGTCCAGGCTGCCATAAGTGGGGAATAAGGAAAATAAAACATGAAACACGGGT
CGGGCAATGTCATGCGGTATTCCGGCTTTGCTTTCCGGCGCGCCTAGGCCGGCCGACATGTACAGAGCTCGAG
GAGCGCCGGAGTATAAATAGAGGCGCTTCGCTACGGAGCGACAATTCAATTCAAACAAGCAAAGTGAACAC
GTCGCTAAGCGAAAGCTAAGCAAATAACAAGCGCAGCTGAACAAGCTAAACAATCTGCAGCCAAGCTGATC
CTCTAGGGTACGCGTAGAGTCGAGAGGCTGTTAAACGATCCACCGGTCGCCACCATGGTGAGCAAGGGC
GAGGAGCTGTTACCAGGGGTGGTGCCATCCTGGTGCAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG
CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA
GCTGCCCGTGCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTACGCCCTACCCCGAC
CACATGAAGCAGCACGACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCA
AGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAG
CTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCA
CAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGA
GGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCATGCTCT
GCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAGCAGGCACAGAAGGC
ATCGCCAGCGCTCTAGGAGCCGCAATCGCAGCCGAAGTCGCAGCAGTGAACGAAAACGCCGTCAACGGAGC
CGAAGTCGCAGCAGTGAACGAAGACGCTACTTGTACAAGTAAAGCGGCCGCGACTCTAGATCATAATCAGCC
ATACCACATTTGTAGAGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATG
AATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAATTT
CACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGGTGGTTGTCCAACTCATCAATGTATCTTACTAGTGGCC
ACGTGGAGGCCCGCCGAATTCGAATGGCCATGGGACGTGCACCTGTGGTAATTATAACCTTGGCGCCTAAGG
ATCGATACTGGGATACCCTGAACCCTCTTCTCCAACACCAAATGCAAAATATATACAGACTCTATATATA
TGCAGACTCTAGCTCTTATTATTTACCCAATTTGCTCAAAAAACAAGATTTGATATCGATTTTATCGATTGCT
TGGAAACGGGGCAAGTTATCGAATATCGGAAACAACTCGATCTGCGTAGACACTAGTCAAGTTCTTCAAGTC
TCTAGCTCTTCTGATATCCCTGCCTTCATACATATGGACGGACGGACGGACGGACATGGCTAGATCGACTCGG
CTATTGATGCTGATCAAGAATATATACTTTATGGGGTCGGAGGTGCTTCATTCTGCCTGTTACATACATTTG
CATTTGCACAAACCTTCTTACCCACTTTCAATGGGTTTCAGAGTATAAAATGAAATTGTTTTGTGCTCATCGA
ATTCGAAATTGTCATTCAAGCAGAAAAATTGAAATCAACTGTGAAATTTCCAATAAATATTAAGCATATACAATT
ATGCATTTGAAGCTATTCAAATTGAAGCTAGAAAATTGAGCATATCTGCGGCACGCCGAAGATGAAATACCCT
TCCTTTCAGATAGGCGCACATTGTCACGCCTCAATTCTCATTTCTCGACCGATTCTTCTTTGATCTATATTA
GATAGTCGTACAATATGAGATGAACTAGTCAATGCTACAGTTAGTCCAGATATCTTGATGAAAACAAAGTTTT
TCGTACAAGAACCTACTTAGCGACCTATCGTTTCTATGCCAGCTATATGATATAGAAAACCGAAGAGATTAC
CTACGACGAATTGCAAGACGTTAGCTAAAATGCTGAGAGACTAGTTCTCAGAGAAACAAACAGATGGATCAG
ATATATATATGGAGATGTTCCCTTCTAGGTGTTACACATTCCATGACAATCTCATAATACCCCGCACAAGCATC
ACTAGTGAATTCGCGGCCGCTGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCTTGA
GTATTCTATAGTGTACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCG

CTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCGTTGCGCTCACTGCCCGCTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAAT
CGGCCAACGCGCGGGGAGAGGCGGTTTGCCTATTGGGCGCTCTTCCGCTTCTCGCTCACTGACTCGCTGCG
CTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG
GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTG
CTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCAAAAAATCGACGCTCAAGTCAGAGGTGGCGAA
ACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCT
GCCGCTTACCGGATACCTGTCCGCTTTCTCCCTTCGGAAGCGTGCGCTTTCTCATAGCTCACGCTGTAGGT
ATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTAGCCCCGACCGCTG
CGCCTTATCCGTAACATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACT
GGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGG
CTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTGTTTGAAGCAGCAGATTACGCGCAGAA
AAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGTCTGACGCTCAGTGGAAACGAAAACCTCACGTTA
AGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTAAATTAATAAAGTAAAAAT
CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGC
GATCTGTCTATTTGTTCCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTAC
CATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCA
GCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGC
CGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTG
GTGTCACGCTCGTCGTTTGGTATGGCTTCATTAGCTCCGGTCCCAACGATCAAGGCGAGTTACATGATCCCC
CATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTTA
TCACTCATGGTTATGGCAGCACTGCATAATTCTTACTGTATGCCATCCGTAAGATGCTTTTCTGTGACTGGT
GAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGCGTCAATACGG
GATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACCTCT
CAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGAC
ACGGAAATGTTGAATACTCATACTCTTCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAG
CGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAAT
ATTTTGTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAAATCGGCAAAATC
CCTATAAATCAAAGAATAGACCGAGATAGGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAA
AGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCAC
CCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAGCACTAAATCGGAACCCTAAAGGGAGCCCCGATTTAG
AGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAG
GGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCACCACACCCGCCGCGTTAATGCGCCGCTACAGGG
CGCGTCCATTGCGCATTAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCT

Supplemental Text S2-2: Annotated pGEM-3XP3.RFP-*tan* sequence (6653bp):

5' homology arm

5' target PAM

Unique portion of 5' target site for reporter excision (restriction sites: BglII, BsiWI, Acc65I)

3XP3 eye-expression Promoter

RFP with α -tubulin 3'UTR

Unique portion of 3' target site for reporter excision (restriction sites: NarI, Bsu36I, ClaI)

3' homology arm

3' target PAM

Restriction sites for homology arm removal: 5' end – ApaI, NcoI; 3' end – SexAI, MluI

```
ATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCAC
GACGTTGTA AACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTGGGCCGACGTCGCATGCTC
CCGGCCGCGATGGCGGCCGCGGAATTCGATCTCTGGATAAGCGTCAGCCTCGCGTTGCGTCCGAGTCGCAGTCG
CAGTCGCCTTCGCTGCTGGCGTCGCAGCGTCGCCATTGCTGTTATTGTTGTTGCTGTGTTTCTATATGCGTTGCC
GCTCCGTTATCGCAGCGACGTCGTCGTCGTCGTCGTCATCAGCTGTCTTCGTGTTGTTGTTGCTTTTTG
CTGAGTGTGTTGTTGTTGCGCACCTCGAATTGCTGCCGCCAAAACATGTCCAATTCAGTTTGAATTCTGCCTTCAAGC
GCTTTAGAACAATCATATAAAAACCTGGCTGAATTATCTATATACAATTTGTTTACTGAAATTCTCTCTCTCTCTC
GTAACCTTCTCTCGCTCTGTCTGTGTGTCGTCGTCGTTGAGTGTGGTAAAGTTGTTAATGAAATTATT
AACTCAGTTGTGTTTTTTTATTATTATTATTTTATCTTGTGAATTTGTTTTGTTTGTGCAAAAATTTGTTGT
AATTCACTTTTCCAGCGAGTTTTTATCCCACTGAATTATTATATATATTTCTGTGCTTATTTTTGTTTCGCTTTC
TATTTACATTTGTGATCGTGGATAAGACGAACCCAAGCAAAAAAAAAAAGAAAAGAAAAGAATAATTGAGAG
CATATTACTGCTCATCTGCTCGAATTCAGAGCCAAATAGCGCGAAGAGCGATAAACGGTCCACAGCACATA
GACCAAAACGAAAGCCCAGGAAGCCATGTGCACCTCCCGGTGGCCACCAATGTCGTCCAGCATGTCCTCCGG
CAAGATCTCGTACGGTACC AATTGAGCTCATAAATTCTGTATAATGTATGCTATACGAAGTTATGTCGACGAATTCG
CGGCCGCGAGCTCGCCCGGGGATCTAATTC AATTAGAGACTAATTC AATTAGAGCTAATTC AATTAGGATCCAAGC
TTATCGATTTGAAACCCTCGACCGCCGAGTATAAATAGAGGCGCTTCGTCTACGGAGCGACAATTC AATTC AAAC
AAGCAAAGTGAACACGTCGCTAAGCGAAAGCTAAGCAAATAAACAAGCGCAGCTGAACAAGCTAAACAATCGGG
CGGCCGCACTAGAGCCGTCGCCACCATGAGGCTTCCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAAGGTTCC
GCATGGAAGGAACGGTCAATGGGCACGAGTTT GAAATAGAAGGCGAAGGAGAGGGGAGGCCATACGAAGGCCA
CAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCCACCAATTTAG
TATGGAAGCAAGGTATATGTCAAGCACCTGCCGACATACCAGACTATAAAAAGCTGTCATTTCTGAAGGATTTA
AATGGGAAAGGGTCACTGAACCTTTGAAGACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCT
GTTTCATCTACAAGGTCAAGTTCATTGGCGTGAACCTTCTTCCGATGGACCTGTTATGCAAAAAGAAGACAATGGG
CTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTGAAAGGAGAGATTATAAGGCTCTGAAGCT
GAAAGACGGTGGTCATTACCTAGTTGAATTC AAAAGTATTTACATGGCAAAGAAGCCTGTGCAGCTACCAGGGTA
CTACTATGTTGACTCCAACTGGATATAACAAGCCACAACGAAGACTATAACAATCGTTGAGCAGTATGAAAAGAACC
GAGGGACGCCACCATCTGTTCTTTAGCGGCCATCGAATTCGAGCTCGCCCACTAAGCGTCGCGCCACTTCAACGC
TCGATGGGAGCGTCATTGGTGGGCGGGGTAACCGTCGAAATCAGTGTTTACGCTTCCAATCGCAACAAAAAATTC
ACTGCAACACTGAAAAGCATACGAAAACGATGAAGATTGTACGAGAAACCATAAAGTATTTTATCCACAAAAGACA
CGTATAGCAGAAAAGCCAAGTTAACTCGGCGATAAGTTGTGTACACAAGAATAAAAATCGGCCAGATTCAAGTGTG
TCAGAAAATAAGAAAACCCCACTATGTTTTCTTTGCCTTTCTTTCTCCAGCGATCATTCAATTCGTGGTGAAAGAA
CGGGGTCAATGCACGGAGTTTCGACTGCGGGAAAGCAGAGCTGCCGTTCACTTCGTCTATAATTAGCGCTTCTAT
```

TTTCCCCGATTCTGGGCCGCTGCTGCGCTTTCCGCTGCTGTTTGTGGCAAGTGTAGCAGCAGGCTGTGCACGCAG
TGTGGCATGCACTTGGCTTTCCACCGTTGGTATCGATTCTCTGGGACGATGAGTCATTCTTTTCGGGGCCACAGCA
TAATCGTTGCCAGCTCACCGAAATGGTGACTTCAATTTCTAACTGCCGTCAAGCATGCGATTGTACATACATACATA
TTTATATATGTACATATTTATGTGACTATGGTAGGTCGATATAATAGCAATCAACGCAAGCAAATGTGTCAGTCTCG
CTTACAGGAACGATTCTATTTAGTAATTTTCGTTGTATAAAAGTAATTATGTATGTATGTAAGCCCCATAAATCTGAA
ACAATTAGGCAAAACCAATGCGAAGCTCTCTGGCGCCTAAGGATCGATGCTAGGCTCTGACTTTTATATAGAATAATC
TTGCCAGATTTGCATCTTTCTCTGAGTGAACCTTGTGTTGATTCTCTCTCTCTCGCTCTCGCCCTCGCTTTCTCGTTCTC
GCTCTCGCTCTCTCTCTTATCAGCAGGTGCAATTACAGCTACATATAGTAATCACATAAAGATAATATGATCCTG
ATAAGCGTTTATACGATTGACTAGTATATCTACTTGTCTTTCTTCTACTTTTTTGGCATTTCGTGACCCAGTTCCT
TAACTTAACTTCACTCAAAACTAACACTCACTCACACTCACACACACACACACACACACAAATATAACAATC
ACTCACTTGTTCGCTTTGAGCCGGAACAAACAAAGAAAAAAGTGGCCAATATCAAGGGCAACAATTGGC
ATGCAGGCCCGACTCCGTCTGGGCGAGCTGCTGGTGGGTTGGGTTGGCAGGTGAGCGGGCGCTGGGCCGGGT
TGCCCGTGGTCTGGACCCACTCTGCTCCGTCGAGCAGTCCCGCGCACATAATTAACAAGCTGGCAACAAGATCT
ATGTCCAACAATAAACCATTATTCATAATGTTGTTTTAGCGTTTCGTTTTGTTGTTGTTGCCCTGTGCATCAGTCGTG
ATCGATTATGGTTCTTTTTGTTGTTGTTGTTGTTGTATATTTGTAAAATATATTTTTGCCAGCAGCAGCATAT
TAGCATAACAATACGTTTAAAGAGGGCAATTTGCAAACGGAGTTCGAATCAAAATTTAGAAAGCCGCTTGCAATTC
AATTCAATATTTCCAATTAATCATAATAATATACATTATATATATATATAAATATATGTGTGTGTGCTTTGC
AGGGTTCGCACTTTTGGCTCGATGATCAAGAATTTCTGATCTTATCGCAGCCACTGAACGAGACCTACTTGGCCGCT
TATGGCACCGCAAAAGGCAGTAAGTACTACTCCACGGACAGTTGCTTCTAGATACGCGTTGGATGCATAGCTT
GAGTATTCTATAGTGTACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTGTTATCCGC
TCACAATTCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCA
CATTAAATGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCA
ACGCGCGGGGAGAGGGCGGTTTGCATATTGGGCGCTCTCCGCTTCTCGCTCACTGACTCGCTGCGCTCGGTCTG
CGGCTGCGGCGAGCGGTATCAGCTCAAAAGGCGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGG
AAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGTGGCGTTTTTCCATA
GGCTCCGCCCCCTGACGAGCATCAGAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTTCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCC
GCCTTTCTCCCTCGGGAAGCGTGGCGTTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGGTGTAGGTCGTTG
CTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCCCGACCCTGCGCCTTATCCGGTAACTATCGTCTTGAG
TCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTA
GGCGGTGCTACAGAGTTCTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCT
CTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCACCGCTGGTAGCGGT
GGTTTTTTTGTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGG
GGTCTGACGCTCAGTGGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTACCTA
GATCCTTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAAT
GCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTTCATCCATAGTTGCCTGACTCCCCGTGCTGTAG
ATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGTGCAATGATACCGCGAGACCCACGCTCACCAGCT
CCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCC
ATCCAGTCTATTAATGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCAGCAACGTTGTTGCCA
TTGCTACAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTCAGCTCCGTTCCCAACGATCAAGGCG
AGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGTCCTCCGATCGTTGTCAGAAGTAAGTTG
GCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTTACTGTGATGCCATCCGTAAGATGCTTTTTCT
GTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGGCCGGCGTCA
ATACGGGATAATACCGCGCCACATAGCAGAACTTAAAGTGTCTCATATTGAAAACGTTCTTCGGGGCGAAAA
CTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACCTGATCTTACGATCTTTT
ACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGGAATAAAGGGCGACACG
GAAATGTTGAATACTCATACTCTTCTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATA
CATATTTGAATGTATTTAGAAAAATAACAAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGATGCG

GTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTAATATTTTGTTAAA
TTCGCGTTAAATTTTGTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAATCCCTTATAAATCAA
AGAATAGACCGAGATAGGGTTGAGTGTGTTCCAGTTTGAACAAGAGTCCACTATTAAGAACGTGGACTCCAA
CGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCACTACGTGAACCATCACCTAATCAAGTTTTTTGGG
GTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCGATTTAGAGCTTGACGGGGAAAGCCGG
CGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCA
CGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGCTGCGC
AACTGTTGGGAAGGGCGATCGGTGCGGGCCTTTCGCT

Supplemental Table S2-1: Oligonucleotides and primers used in reagent construction and molecular screening: All base-pairing nucleotides are in capital letters. Tailing sequences added to primers are in lowercase letters.

Pair #	Purpose	Forward (description)	Reverse (description)
1	Screen for plasmid backbone integration on the 5' end of the donor. (282bp band if backbone is present)	GATGTGCTGCAAGGCGATTA (In pGEM-WingGFP-tan plasmid backbone, 5' of homology arm)	GAGCAGGCAACGCATATAGAAA C (Within the pGEM-WingGFP-tan 5' homology arm)
2	Screen for plasmid backbone integration on the 3' end of the donor. (481bp band if backbone is present)	GCGACCTATCGTTCCTATGC (Within the pGEM-WingGFP-tan 3' homology arm)	CAGCTGGCACGACAGGTTTC (In pGEM-WingGFP-tan plasmid backbone, 3' of homology arm)
3	Confirm wing-GFP reporter insertion at correct locus on the 5' side. (1541bp band if correctly integrated)	AAACGAACCGCAACTGATATTGAAC (In <i>D. americana tan</i> transgene DNA outside the 5' homology arm. Represented as primer X in Figure 2-2.)	GTCAATTTCCACTGTCCCGATTG (In the pGEM-WingGFP-tan reporter sequence, specifically the pupal wing enhancer from the <i>D. melanogaster yellow</i> gene)
4	Confirm wing-GFP reporter insertion at correct locus on the 3' side. (1163bp band if correctly integrated)	TGGTTTGTCCAACTCATCAA (In the pGEM-WingGFP-tan reporter sequence, specifically in GFP)	AATATAGAGCGCAGCGGCTGTT (In <i>D. americana tan</i> transgene DNA outside the 3' homology arm. Represented as primer Y in Figure 2-2.)
5	Amplify 5' homology arm for pGEM-WingGFP-tan tailed with Gibson assembly overlaps	catggcgccgcccgaattcgatCTCTGGATA AGCGTCAGCCTC (Represented as primer A in Figure 2-2.)	gatgagtatggatggtaccgtacgatct TGCCGGAGGACATGCTGGAC

6	Amplify 3' homology arm for pGEM-WingGFP-tan tailed with Gibson assembly overlaps	ttataaccctggcgcctaaggatcgatACTTGG GGATACCCTGAACC	gccgcaattcactagtgatGCTTGTGC GGGGTATTATGAG (Represented as primer B in Figure 2-2)
7	Amplify wing-expressing GFP reporter tailed with Gibson assembly overlaps	agatctcgtaggtaccATCCATACTCATCAATG TATCTTATAGCTTATCGTTACGTGTA	agtatgatccttagcgccAGGGTTATAA TTACCACAGGTCG
8	Amplify <i>tan</i> sequence to replace wing-GFP reporter in second stage of allele replacement	catggcgccgcggaattcgatCTCTGGATA AGCGTCAGCCTC (Same forward primer as in Pair #5 from this table. Represented as primer A in Figure 2-2.)	gccgcaattcactagtgatGCTTGTGC GGGGTATTATGAG (Same reverse primer as in Pair #6 from this table. Represented as primer B in Figure 2-2.)
9	Spanning PCR across edited <i>tan</i> locus after second replacement step (~2.5kb band if repaired correctly)	AAACGAACCGCAACTGATATTGAAC (Same forward primer as in Pair #3 from this table. Represented as primer X in Figure 2-2.)	AATATAGAGCGCAGCGGCTGTT (Same reverse primer as in pair #4 from this table. Represented as primer Y in Figure 2-2.)
10	Amplify new 3' homology arm which in the modified pGEM-WingGFP-tan before the construction of pGEM-3XP3.RFP-tan (cloned into pGEM-WingGFP-tan using Bsu36I and MluI)	atcctgcctaaggatcGATGCTAGGTCTGAC TTTTATATAGAATAATC (Tailed with Bsu36I and ClaI restriction sites.)	gatacaacgcgtatctGAGAAGCAAC CTGGTC (Tailed with MluI restriction sites.)
11	Amplify 3XP3-RFP reporter	ttatggcgcgCCAACGTGTCGGTACCAATT G	atcgatccttagcgccAGAGAGCTTC GCATGGTTTTGC

	out of M{3XP3-RFP.attP}ZH-51C landing site to assemble pGEM-3XP3.RFP-tan	(Contains Acc65I restriction site in the primer.)	(Tailed with NarI and Bsu36I restriction sites.)
12	Oligo-nucleotide pair for cloning sgRNA expression plasmid pCFD3 targeting t5 site (see Figure 2-2)	gtcGCGACGGGGCAATATCTTGC (Sense oligo. which matches t5 target site. Note that the sgRNA does not include the PAM nucleotides. "gtc" tail is for cloning into the BbsI restriction site in pCFD3. Instructions for pCFD3 cloning at: http://www.crisprflydesign.org/wp-content/uploads/2014/05/Cloning-with-pCFD3.pdf)	aaacGCAAGATATTGCCCCGTCG (Antisense oligo. to anneal with sense oligo to form insert for cloning. "aaac" tail is for cloning into the BbsI restriction site in pCFD3.)
13	Oligo-nucleotide pair for cloning sgRNA expression plasmid pCFD3 targeting t3 site (see Figure 2-2)	gtcG TTCAGGGTATCCCCAAGTC (See description for Pair #12)	aaacGACTTGGGGATACCCTGAA (See description for Pair #12)
14	Oligo-nucleotide pair for cloning sgRNA expression plasmid pCFD3 targeting t5re site (see Figure 2-2)	gtcGGTACCGTACGAGATCTTGC (See description for Pair #12)	aaacGCAAGATCTCGTACGGTAC (See description for Pair #12)
15	Oligo-nucleotide pair for cloning sgRNA expression plasmid pCFD3 targeting t3re site (see Figure 2-2)	gtcGGCGCCTAAGGATCGATACT (See description for Pair #12)	aaacAGTATCGATCCTTAGGCGC (See description for Pair #12)
Seq1	Sanger sequencing 5' PolyT	CTCTGGATAAGCGTCAGCCTCG	

Seq2	Sanger sequencing 5' PolyT	CGCCCATTGCTGTTATTGTT	
Seq3	Sanger sequencing 5' target sites	ATTACTGCTCATCCTGCTCG	
Seq4	Sanger sequencing 3' PolyT	TCACTATGAGGTCGGCTTCG	
Seq5	Sanger sequencing 3' PolyT and 3' target sites		TTGACTAGTGTCTACGCAGATCG

Embryos injected	Injected adults crossed	Founders ¹	Homozygous edited lines established from progeny of founders	Proportion of founders' progeny with wing GFP expression	Homozygous lines with carrying correct ² repair events
1220	150	6	43	2.5%-25.4%	13

1: In this experiment, founders are defined as injected adults who produced progeny with pupal wing GFP expression.

2: Correct repair events were identified first by the PCR screens described in the results section. All 13 lines that passed the PCR screen were later verified to carry correctly repaired sequence by Sanger sequencing.

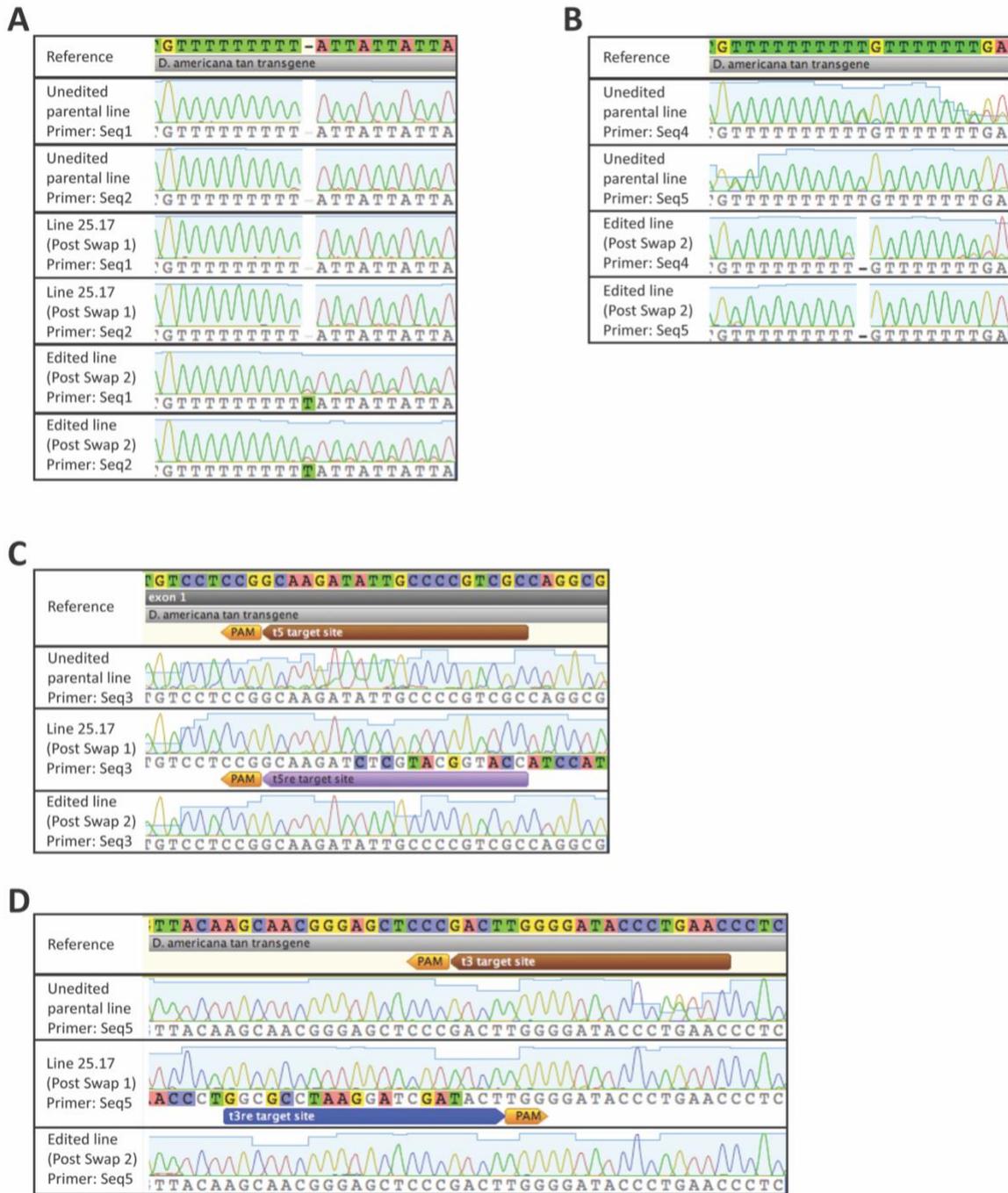
Supplemental Table S2-2: Genome editing rates in the first stage of the allele swap

Embryos injected	Injected adults crossed	Founders ¹	Homozygous edited lines established from progeny of founders	Proportion of founders' progeny that lost wing GFP expression	Homozygous lines with carrying correct ² repair events
1361	179	5	10	0.68%-1.1%	1

1: In this experiment, founders are defined as injected adults who produced progeny that did not express GFP in the pupal wings.

2: Correct repair events were identified first by the PCR screen described in the results section. The single line that passed the PCR screen was later verified to have a correctly repaired sequence by Sanger sequencing.

Supplemental Table S2-3: Genome editing rates in the second stage of the allele swap



Supplemental Figure S2-1: Chromatograms showing the sequence of the edited sites before and after each stage of the editing process. In all panels, the reference sequence in the top row represents the unedited *D. americana tan* transgene that was targeted for genome editing in this study. The sequencing primers are listed to the left of the chromatograms, and all primer sequences are provided in Supplemental Table S2-1. Panels (A) and (B) show chromatograms of the edited homopolymer runs in the 5' homology arm (A) and in first intron (B) of the *D. americana tan* transgene throughout the allele swap process. Sequencing data for homopolymer runs are shown for two different primers per DNA

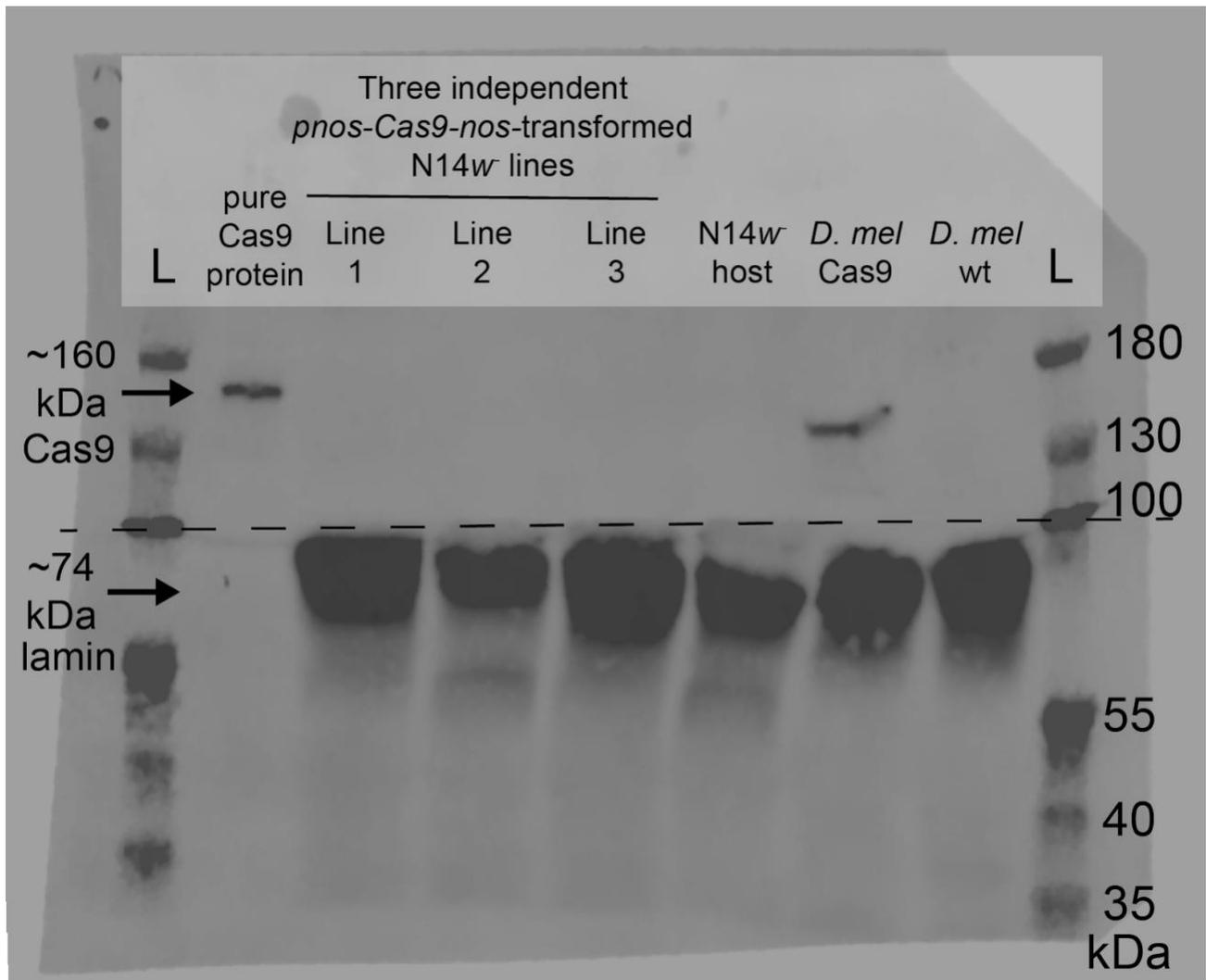
template. Panels (C) and (D) show chromatograms of the 5' target sites and 3' target sites, respectively, across the stages of genome editing. The PAM sequences and sgRNA target sites are labeled.

Appendix B

Supplemental Data for Chapter 3



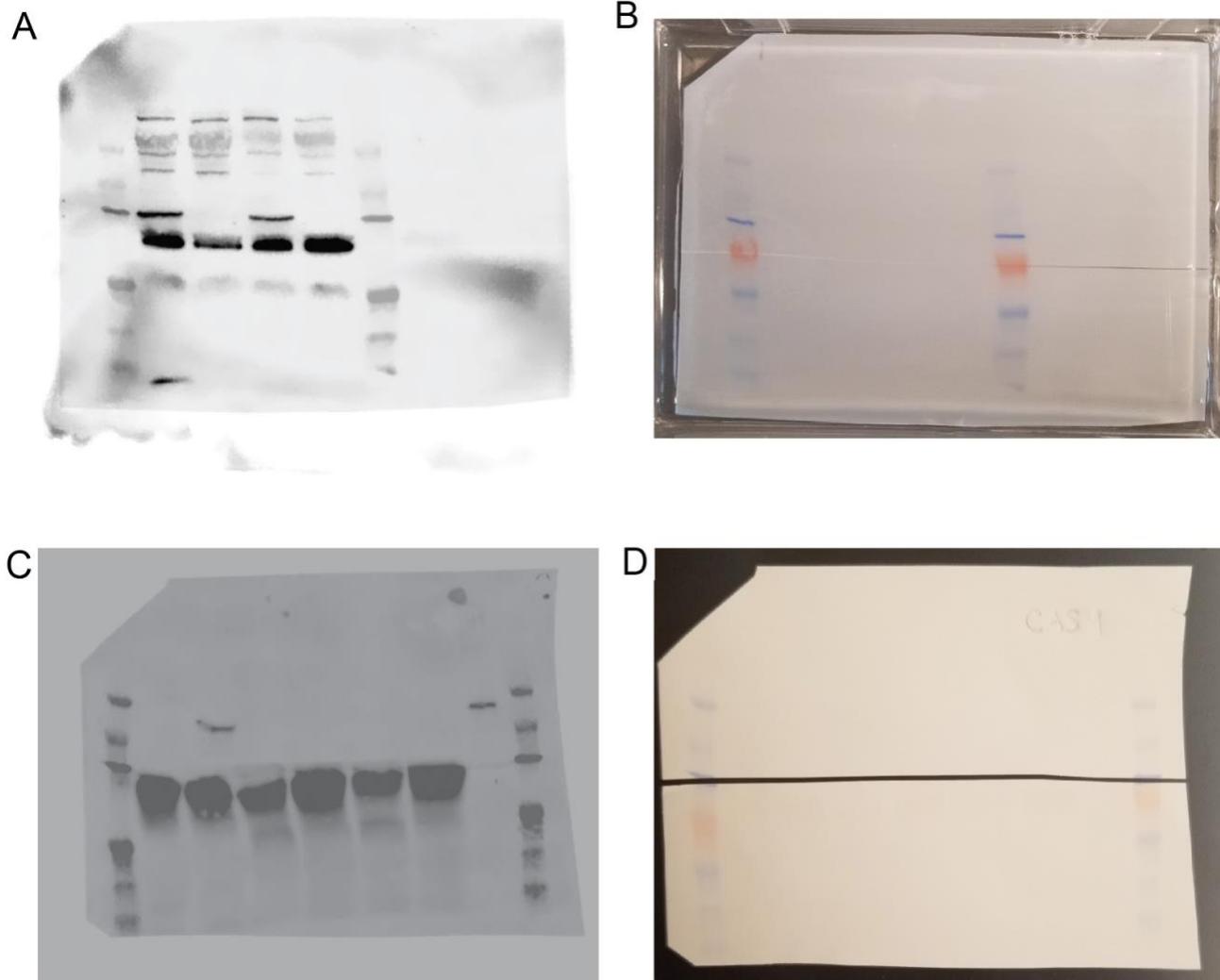
Supplemental Figure S3-1. Mutations in the *D. novamexicana* and *D. americana white* genes cause white-eyed phenotype. Photograph shows *D. americana white* mutant adult male (left) alongside *D. novamexicana white* mutant adult male (right).



Supplemental Figure S3-2. *nanos* promoter failed to drive expression of Cas9 in transgenic *D. novamexicana*. Western blotting showed that three independent insertions of a transgene expected to express Cas9 protein in the germline under the control of the *nanos* promoter failed to produce detectable levels of Cas9. From left to right, lanes contained a sample of pure Cas9 protein obtained commercially (a positive control) and then total protein extracts from ovaries of sexually mature adult female flies from three lines of *D. novamexicana* with independent insertions of the piggyBac transgene (lines 1-3), the *D. novamexicana* N14 *white* mutant (*w*) host line transgenes were injected into (a negative control), a line of *D. melanogaster* expressing Cas9 in the germline (a second positive control), and a wild-type (wt) of line of *D. melanogaster* (a second negative control). "L" represents the ladder, or molecular weight marker (PageRuler prestained protein ladder). The dotted black line shows where the membrane was cut prior to incubation with primary antibodies during the western blotting procedure; the top half was incubated with anti-Cas9 antibodies whereas the bottom half was incubated with anti-Lamin antibodies. The two halves were realigned by hand for imaging, using the shape of the cut and the ladder staining as a guide. Relative intensity of the protein detected with the antibody against Lamin estimate the relative amounts of total protein loaded per lane. An un-annotated image of this blot is shown in Supplemental Figure S3-4.



Supplemental Figure S3-3. Mutation in the *D. novamexicana* *yellow* gene causes visible changes in body pigmentation. Photograph shows *D. novamexicana* *white* mutant adult male (left) alongside *D. novamexicana* *white, yellow* mutant adult male (right). Consistent with *yellow* null phenotypes in other species, *D. novamexicana* individuals identified as *yellow* mutants displayed a complete lack of black pigmentation on body, wings, and bristles. Phenotypically *yellow* males were observed only in crosses where the female N14 *nos-Cas9-nos* transgenic parent was injected with Cas9 protein along with sgRNAs targeting conserved sites in *yellow* exon 1.



Supplemental Figure S3-4. Uncropped and un-annotated western blot images and membranes. (A-B) Images show combined enhanced chemiluminescence (ECL) and 700nm ladder channels **(A)** and photograph of the blocked and cut membrane **(B)** from Ebony western blot in Figure 3-3B. **(C-D)** Images show combined enhanced chemiluminescence (ECL) and 700nm ladder channels **(C)** and photograph of the cut membrane **(D)** from Cas9 western blot in Supplemental Figure S3-2.

Supplemental Table S3-1. Oligonucleotides used to generate sgRNAs for CRISPR/Cas9 genome editing and *nos-Cas9-nos* transgene.

Name	Purpose	Sequence
IVT_Common_Rev	Common reverse primer for generating all DNA templates for in-vitro transcription of sgRNAs	AAAAGCACCGACTCGGTGCCACTT TTTCAAGTTGATAACGGACTAGCCT TATTTTAACTTGCTATTTCTAGCTCT AAAAC
IVT_e_ex2A_Fw	Forward primer for generating DNA template for in-vitro transcription of sgRNA targeting one site in <i>ebony</i> exon 2. Target site in bold.	GAAATTAATACGACTCACTATAGGC TGCGCCATGCCGACAAGGG TTTTA GAGCTAGAAATAGC
IVT_e_ex2B_Fw		GAAATTAATACGACTCACTATAGGC TCATTTATCAGCAGGAACG TTTTA GAGCTAGAAATAGC
IVT_e_ex2C_Fw		GAAATTAATACGACTCACTATAGGT GCCGCCCGTCTCCTTGTGGG TTTT AGAGCTAGAAATAGC
IVT_e_ex2D_Fw		GAAATTAATACGACTCACTATAGGC AGCGATGGGGACTTCATCGG TTTT AGAGCTAGAAATAGC
IVT_e_ex2E_Fw		GAAATTAATACGACTCACTATAGGG TTTGCATGCAACCGTCGGG TTTTA GAGCTAGAAATAGC

IVT_w_ex2_Fw	Forward primer for generating DNA template for in-vitro transcription of sgRNA targeting one site in either <i>white</i> exon 2 or 3. Target site in bold.	GAAATTAATACGACTCACTATAGGT CCACATTGTGCCACGCATGTTTTA GAGCTAGAAATAGC
IVT_w_ex3_Fw		GAAATTAATACGACTCACTATAGGA CCACGCTGCTGAATGCCCGTTTTA GAGCTAGAAATAGC
IVT_yA_Fw	Forward primer for generating DNA template for in-vitro transcription of sgRNA targeting one site in <i>yellow</i> exon 1. Target site in bold.	GAAATTAATACGACTCACTATAGGA AGGAGCAGGCGATCGCCAGGTTT TAGAGCTAGAAATAGC
IVT_yB_Fw		GAAATTAATACGACTCACTATAGGC CGCAGAACGGCCTGCCCGTGTTTT AGAGCTAGAAATAGC
IVT_yC_Fw		GAAATTAATACGACTCACTATAGGT CTTTGTCACAGTGCCGCGGTTTT AGAGCTAGAAATAGC
nosCas9_GA_Fw	Amplify <i>nos-Cas9-nos</i> transgene out of <i>pnos-Cas9-nos</i> (Addgene #62208) with tailed primers to assemble into pBac{3XP3-ECFPafm}.	TTCGAATGGCCATGGGACGTCGAC CGGATTTCACTGGAAGCTAGGCTAG
nosCas9_GA_Rev		ATATAGGGCCCGGGTTATAATTAC CCGAGACCGTGACCTACATCG

Supplemental File S3-1. Raw data measuring CHC abundance Download at:

https://deepblue.lib.umich.edu/data/concern/data_sets/pc289j48s

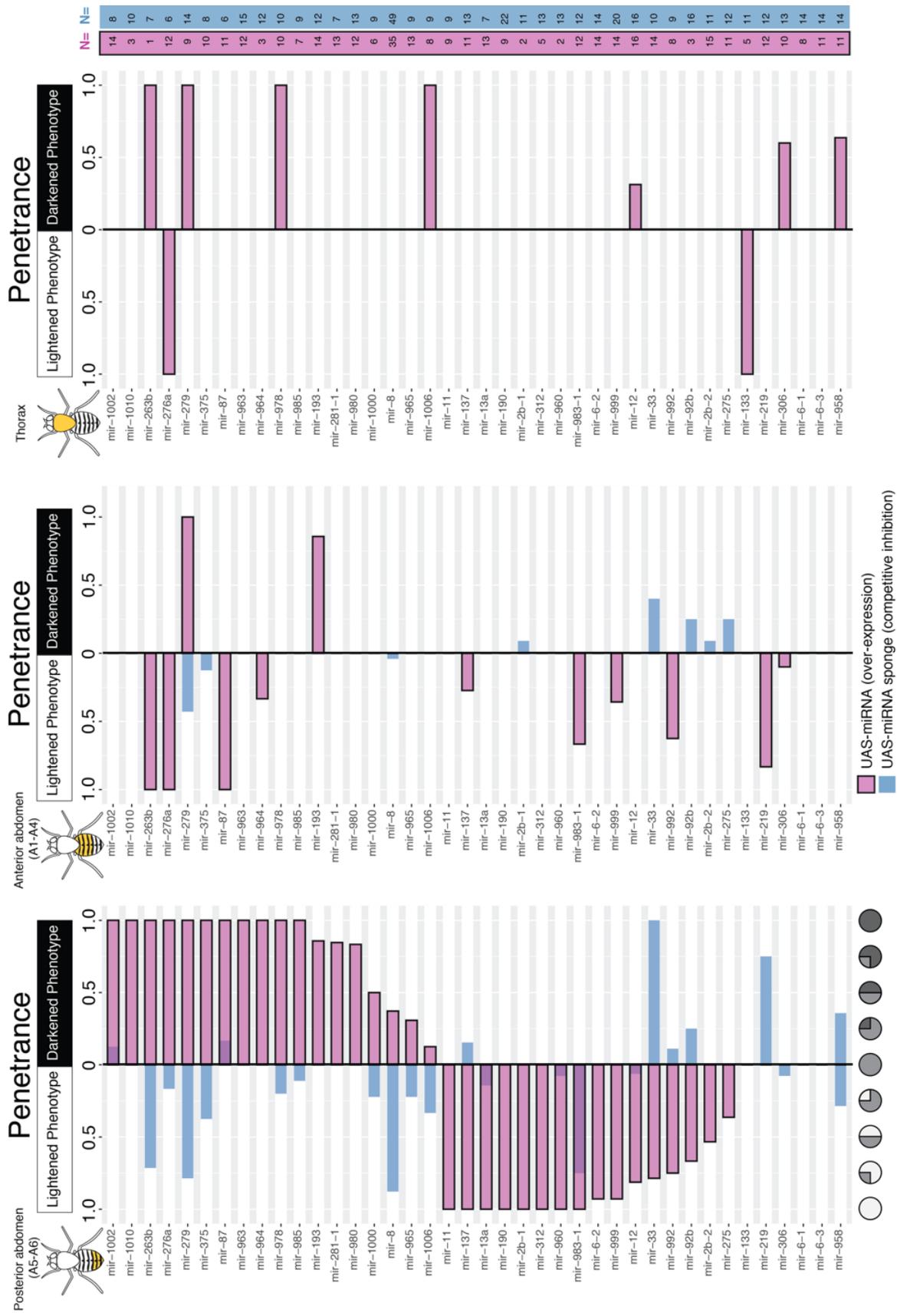
Supplemental File S3-2. R code used for analyzing CHC data Download at:

https://deepblue.lib.umich.edu/data/concern/data_sets/pc289j48s

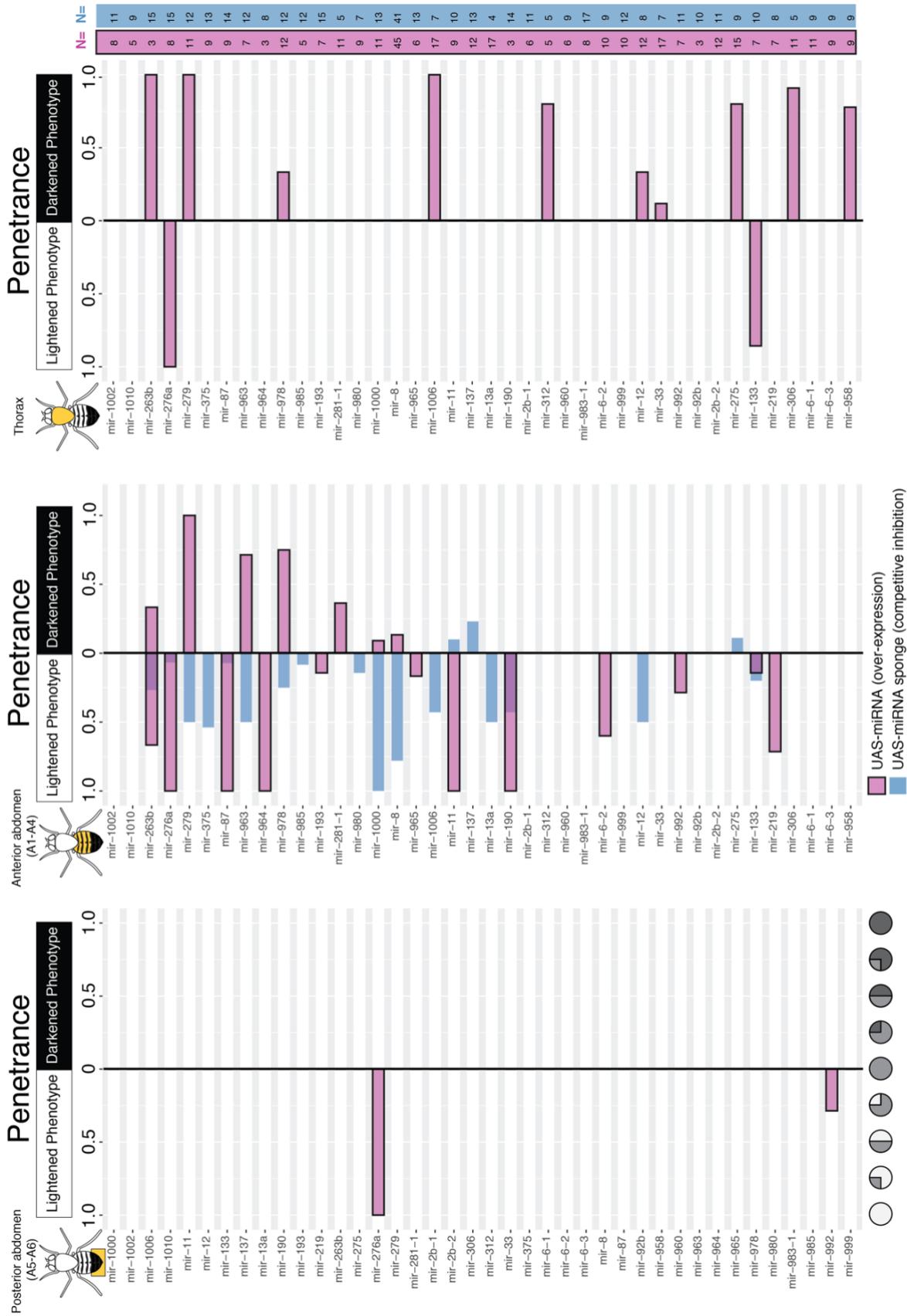
Appendix C

Supplemental Data for Chapter 4

Supplemental Figure S4-1: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen, male data, competitive inhibition crosses at 28°C. From left to right, the plots represent penetrance of phenotypes observed in the posterior abdomen, anterior abdomen, and thorax of male flies. Penetrance of phenotypes range from 0 (no flies observed display plotted phenotype) to 1 (all flies observed display plotted phenotype). Penetrance of lightening phenotypes are plotted extending to the left from the “0” axis, while penetrance of darkening phenotypes are plotted extending to the right. Competitive inhibition data is represented by blue bars without borders, while over-expression data is represented by pink bars with black borders. Numbers of flies phenotyped for each penetrance value are listed on the right of the figure under “N=”, and are color coded in the same manner as penetrance data. All competitive inhibition data displayed was collected from crosses set at 28°C. All lines and genotypes are the same as those described in Figure 4-3.



Supplemental Figure S4-2: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen, female data, competitive inhibition crosses at 18°C. From left to right, the plots represent penetrance of phenotypes observed in the posterior abdomen, anterior abdomen, and thorax of female flies. Penetrance of phenotypes range from 0 (no flies observed display plotted phenotype) to 1 (all flies observed display plotted phenotype). Penetrance of lightening phenotypes are plotted extending to the left from the “0” axis, while penetrance of darkening phenotypes are plotted extending to the right. Competitive inhibition data is represented by blue bars without borders, while over-expression data is represented by pink bars with black borders. Numbers of flies phenotyped for each penetrance value are listed on the right of the figure under “N=”, and are color coded in the same manner as penetrance data. All competitive inhibition data displayed was collected from crosses set at 18°C. All lines and genotypes are the same as those described in Figure 4-3.



Supplemental Figure S4-3: Penetrance of pigmentation phenotypes for all miRNAs included in competitive inhibition screen, male data, competitive inhibition crosses at 18°C. From left to right, the plots represent penetrance of phenotypes observed in the posterior abdomen, anterior abdomen, and thorax of male flies. Penetrance of phenotypes range from 0 (no flies observed display plotted phenotype) to 1 (all flies observed display plotted phenotype). Penetrance of lightening phenotypes are plotted extending to the left from the “0” axis, while penetrance of darkening phenotypes are plotted extending to the right. Competitive inhibition data is represented by blue bars without borders, while over-expression data is represented by pink bars with black borders. Numbers of flies phenotyped for each penetrance value are listed on the right of the figure under “N=”, and are color coded in the same manner as penetrance data. All competitive inhibition data displayed was collected from crosses set at 18°C. All lines and genotypes are the same as those described in Figure 4-3.

Supplemental Table S4-1: Phenotypes and notes on all individual flies collected in miRNA overexpression screen. Each row contains the gene under UAS control (“UAS_line”) for each cross, the miRNA associated with the line, date of collection, date of phenotyping, sex, balancers present, notes on phenotypes observed, classification of phenotype in posterior abdomen, anterior abdomen, and thorax (“l” for lightened, “d” for darkened, blank if unaffected), presence of visible developmental defects in abdomen, thorax, bristles, or wings (1 = defects observed), whether the defects obscure the pigmentation phenotype (1= slightly obscured, 2= completely obscured, 0=not obscured), and whether observed defects occurred in the same body segment as the observed pigmentation effects or different body segment (“s” = same, “d” = different). Download at: https://deepblue.lib.umich.edu/data/concern/data_sets/pc289j48s

Supplemental Table S4-2: Phenotypes and notes on all individual flies collected in miRNA competitive inhibition screen. Each row contains the miRNA sponge or control (scrambled) sponge under UAS control (“UAS_line”) for each cross, the miRNA(s) associated with the line, the rearing temperature, date of collection, date of phenotyping, sex, balancers present, copy number of sponge transgenes inherited (default 2 copies, one on each of 2nd and 3rd chromosome, “1” indicates single copy of sponge transgene), notes on pigmentation, phenotypes observed, classification of phenotype in posterior abdomen, anterior abdomen, and thorax (“l” for lightened, “d” for darkened, blank if unaffected), presence of visible developmental defects in abdomen, thorax, bristles, or wings (1 = defects observed), notes on non-pigment phenotypes, whether the defects obscure the pigmentation phenotype (1= slightly obscured, 2= completely obscured, 0=not obscured), and whether observed defects occurred in the same body segment as the observed pigmentation effects or different body segment (“s” = same, “d” = different). Download at: https://deepblue.lib.umich.edu/data/concern/data_sets/pc289j48s

Supplemental Table S4-3: Summary data from overexpression and competitive inhibition screens.

Includes: Fly strain identification numbers, miRNA annotation confidence, counts of flies collected, statistical tests of viability, brief descriptions of pigmentation phenotypes, and brief descriptions of non-pigmentation phenotypes/defects observed. Sheet one contains all data. Sheet two contains descriptions of the contents of columns in sheet one. Download at:

https://deepblue.lib.umich.edu/data/concern/data_sets/pc289j48s

Supplemental Table S4-4. List of pigmentation genes. For each gene, this table lists: name, annotation symbol (“CG” number), categorization as transcription factor (Y = annotated as transcription factor, N = not annotated as transcription factor), pigmentation role (“darkens” if the gene has been experimentally demonstrated as either necessary or sufficient for the development of dark pigmentation; “lightens” if experimental evidence shows that it is necessary to prevent the development of dark pigments and/or sufficient to lighten pigmentation where misexpressed), Loss-of-function effect on female A6 melanization, citations supporting the gene’s role in pigmentation. Citation list: [1] = (Kalay et al., 2016), [2] = (Rogers et al., 2014), [3] = (Dembeck et al., 2015b), [4] = (Wright, 1987), [5] = (Wittkopp et al., 2003), [6] = (Kopp and Duncan, 1997), [7] = (Riedel et al., 2011), [8] = (Dewey et al., 2004), [9] = (Wakabayashi-Ito et al., 2011), [10] = (Norgate et al., 2006), [11] = (Shakhmantsir et al., 2014), [12] = (Baker and Truman, 2002), [13] = (Sekine et al., 2011).

Gene	Annotation symbol	Transcription factor	Pigmentation role	Loss-of-function effect on female A6 melanization	Citation(s)
Abd-A	CG10325	Y	darkens	Less melanized	[2]
Abd-B	CG11648	Y	context-dependent	Less melanized	[1],[2]
Akt1	CG4006	N	darkens	Unknown	[11]
Ask1	CG4720	N	darkens	None	[13]
ato	CG7508	Y	context-dependent	Less melanized	[1]
ATP7	CG1886	N	lightens	Unknown	[10]
b	CG7811	N	lightens	More melanized	[4],[5]
bab1	CG9097	Y	lightens	More melanized	[1],[2]
bab2	CG9102	Y	lightens	More melanized	[2]
BEAF-32	CG10159	Y	context-dependent	Less melanized	[1]
Br140	CG1845	Y	darkens	Less melanized	[1]
brm	CG5942	Y	darkens	None	[1]
btn	CG5264	Y	darkens	Less melanized	[3]
Burs	CG13419	N	darkens	Unknown	[8]
C15	CG7937	Y	darkens	Less melanized	[1]
CG10348	CG10348	Y	darkens	Less melanized	[2]

CG30020	CG30020	Y	darkens	Less melanized	[1]
Chrac-14	CG13399	Y	darkens	Less melanized	[1]
crol	CG14938	Y	context-dependent	Less melanized	[2]
da	CG5102	Y	darkens	Less melanized	[2]
dalao	CG7055	Y	darkens	Less melanized	[2]
dally	CG4974	N	darkens	Less melanized	[3]
DAT	CG3318	Y	lightens	Unknown	[4],[5]
Ddc	CG10697	N	darkens	Less melanized	[4],[5]
dsb	CG1887	N	darkens	Less melanized	[3]
dsx	CG11094	Y	context-dependent	More melanized	[1],[2]
e	CG3331	N	lightens	More melanized	[4],[5]
Efa6	CG31158	N	lightens	More melanized	[3]
Eip74EF	CG32180	Y	lightens	More melanized	[2]
Eip78C	CG18023	Y	context-dependent	Less melanized	[1]
exd	CG8933	Y	lightens	None	[2]
Fili	CG34368	N	lightens	More melanized	[3]
frm	CG10625	N	darkens	Less melanized	[3]
fru	CG14307	Y	darkens	Less melanized	[1]
Glut1	CG43946	N	lightens	More melanized	[3]
grh	CG42311	Y	lightens	More melanized	[2]
Gug	CG6964	Y	darkens	Less melanized	[2]
hb	CG9786	Y	darkens	Less melanized	[1]
Hesr	CG5927	Y	darkens	Less melanized	[1]
Hr38	CG1864	Y	darkens	Less melanized	[1]
Hr4	CG16902	Y	darkens	Less melanized	[2]

Hr46	CG33183	Y	context-dependent	Less melanized	[1]
Hr78	CG7199	Y	context-dependent	Less melanized	[1]
hth	CG17117	Y	lightens	None	[2]
jing	CG9397	Y	darkens	Less melanized	[1],[2]
kay	CG33956	Y	darkens	Less melanized	[3]
Kcmf1	CG11984	Y	lightens	More melanized	[1]
kkv	CG2666	N	darkens	Less melanized	[3]
klar	CG17046	N	lightens	More melanized	[3]
Klp61F	CG9191	N	lightens	More melanized	[3]
lab	CG1264	Y	darkens	Less melanized	[1]
Lim3	CG10699	Y	darkens	Less melanized	[1]
lmd	CG4677	Y	lightens	None	[2]
loco	CG5248	N	lightens	More melanized	[3]
M1BP	CG9797	Y	darkens	Less melanized	[2]
Mad	CG12399	Y	darkens	Less melanized	[2]
MBD-like	CG8208	Y	lightens	None	[2]
MBD-R2	CG10042	Y	darkens	Less melanized	[2]
Met	CG1705	Y	context-dependent	Less melanized	[1]
mgl	CG42611	N	lightens	Unknown	[7]
Mi-2	CG8103	Y	lightens	More melanized	[2]
noc	CG4491	Y	darkens	Less melanized	[1]
omb	CG3578	Y	darkens	Less melanized	[4],[6]
osa	CG7467	Y	darkens	Less melanized	[2]
pdm3	CG11641	Y	lightens	More melanized	[2]
pita	CG3941	Y	darkens	Less melanized	[2]

ple	CG10118	N	darkens	Less melanized	[4],[5]
pns	CG7852	N	darkens	Less melanized	[3]
pnt	CG17077	Y	context-dependent	Less melanized	[1]
rk	CG8930	N	darkens	Less melanized	[12]
ru	CG1214	N	darkens	Less melanized	[3]
sbb	CG5580	Y	darkens	Less melanized	[2]
scrt	CG1130	Y	darkens	Less melanized	[2]
sd	CG8544	Y	darkens	Less melanized	[1]
sima	CG7951	Y	darkens	Less melanized	[1]
sinu	CG10624	N	darkens	Less melanized	[3]
sox102F	CG11153	Y	darkens	Less melanized	[1],[2]
SoxN	CG18024	Y	darkens	Less melanized	[1]
Ssrp	CG4817	Y	darkens	Less melanized	[2]
Su(var)2-10	CG8068	Y	darkens	Less melanized	[2]
Su(z)12	CG8013	Y	context-dependent	Less melanized	[1]
Suchb	CG10622	N	darkens	Less melanized	[3]
t	CG12120	N	darkens	Less melanized	[4],[5]
tai	CG13109	Y	lightens	More melanized	[2]
tfc	CG9134	N	darkens	Less melanized	[3]
Tip60	CG6121	Y	lightens	More melanized	[1]
Torsin	CG3024	N	darkens	Unknown	[9]
tx	CG5441	Y	darkens	Less melanized	[2]
unpg	CG1650	Y	darkens	Less melanized	[2]
ush	CG2762	Y	context-dependent	Less melanized	[1]
vfl	CG12701	Y	lightens	More melanized	[2]

vvl	CG10037	Y	darkens	Less melanized	[1]
y	CG3757	N	darkens	Less melanized	[4],[5]