With Flying Colors: Using *Drosophila* Pigmentation to Study How Differences in Traits Arise.

By

Alisha Victoria John

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
2017

Doctoral Committee:

Professor Patricia J. Wittkopp, Chair
Associate Professor Scott E. Barolo
Associate Professor Gyorgyi Csankovszki
Assistant Professor Monica Dus

Alisha V. John

alishav@umich.edu

ORCID iD: 0000-0002-7610-8494

Dedication

To all those who made this journey possible -
my family for their unwavering support and unconditional love,
my friends for their welcomed distractions and encouragement,
JG for believing in me and standing with me through all of life's adventures,
for all young women who dream of exploring STEM fields,
and for the younger version of myself – that little girl with a fierce love of science and unending curiosity, may she live forever in my soul and fuel my future endeavors.

Acknowledgements

Throughout this journey, I was surrounded by many individuals who helped make this achievement not only possible, but also enjoyable and worthwhile. They helped celebrate the triumphs and made the challenges tolerable. First, I will forever be thankful for the guidance, support, and inspiration offered by my advisor, Trisha Wittkopp. Over the last five years, she has encouraged me to be the best scientist I can be – sometimes directly and sometimes indirectly by being a brilliant, eloquent, and creative scientist. Her influence and support in my life, both personally and professionally, is undeniable and I will forever be thankful to have spent these years working with such an incredible mentor.

I could not imagine a better environment to have completed the work presented than in the Wittkopp Lab. I was constantly surrounded by immensely talented and helpful scientists, creating an environment in which everyone could thrive and work together to encourage the production of great scientific study. I am also grateful for all the coworkers who have become friends – this journey would not have been the same without our lab adventures. I would give specific thanks, but it would simply be a list of all of those who have been part of the Wittkopp Lab since 2011.

I would also like to acknowledge the Program in Biomedical Sciences (PIBS) and the department of Molecular, Cellular, & Developmental Biology (MCDB). PIBS allowed me the freedom to explore my scientific interests during my first year of graduate school; without that opportunity, I would not have been able to discover my interest in gene regulation and evolution

nor find the Wittkopp Lab. In addition, PIBS has offered incredible support throughout my graduate studies – both financially by supporting my travel to scientific conferences and professionally by offering opportunities for career exploration, professional development, and networking. After my first year, MCDB welcomed me into their department and I am appreciative for the sense of community and support I found there. I would especially like to thank my fellow graduate students, specifically those on the MCDB Graduate Student Council, for the sense of comradery it brought.

One of my biggest passion projects outside of research these past years has been science communication and outreach. These activities, and the individuals involved in them, impacted my development as a scientist greatly. I am better able to communicate my enthusiasm for science and inform others after these experiences. Specifically, I'd like to thank the founding members of MiSciWriters – Ada Hagan, Bryan Moyers, Irene Park, Kevin Boehnke, and Shweta Ramdas; working with these individuals sharpened my writing and editing skills and encouraged me to explore scientific interests outside of the work presented here. I'd also like to acknowledge the following opportunities for their influence on my personal and professional development: the Portal to the Public Science Communication Fellowship at the Detroit Zoo, FEMMES at the University of Michigan, the WISE GISE Summer Camp Human Genetics Focus Group (specifically Anat Belasen for being an amazing person to develop a curriculum with), and my cohort at the National Association of Interpretation Certified Interpretive Guide training course at the Brookfield Zoo in Chicago, IL.

While the importance of the scientific support I received is undeniable, the personal support is equally important. In this regard, my family and friends must be acknowledged. To my mom, who encouraged me to complete one more rotation (which happened to be in the Wittkopp

Lab) before deciding that my place in the scientific community wasn't going to be found at the University of Michigan. To my dad, for always waiting on the sidelines in case I wanted or needed additional support. To my stepdad, for planting the seed of attending graduate school in my mind early in life and for always believing I could accomplish anything I set my mind to. To my sister, for being an inspiration for how to keep a full schedule and still be successful. To my nephews, for inspiring me to not only tell them, but also show them that science is cool and that with hard work they can accomplish their dreams. To my entire family, for their unconditional love, their unwavering support, and their willingness to drop everything if/when I need it. To those friends who have become family, you give my life color; thank you for the laughter and encouragement through it all. And lastly, to JG, for always understanding, always accepting, always appreciating, always believing – this journey would have been far less enjoyable without you to share it with.

Finally, to all of those not specifically mentioned who played a role in this great accomplishment, thank you.

Table of Contents

List of Tables

List of Figures

Abstract

Understanding how genotypes influence the production of novel phenotypes and contribute to phenotypic diversity is a fundamental goal in biology. When testing for functional divergence in alleles contributing to phenotypic divergence, researchers often utilize transgenic animals to examine the effects of divergent alleles. Here, I test the commonly held assumption that using of a single, defined genomic location to test for functional divergence between alleles effectively controls for genomic position effects. I find that the relative difference detected between alleles varied across genomic locations, including a single genomic location which allowed expression sufficient to rescue a mutant phenotype, but that failed to detect functional divergence between alleles that was present at other genomic locations. Taking these results into consideration, I used transgenic *Drosophila melanogaster* flies to investigate the role of noncoding DNA sequences in *D. americana* and *D. novamexicana tan* in phenotypic divergence in pigmentation between these species. I found that the 5' half on *tan* intron 1 from *D. novamexicana* in an otherwise *D. americana tan* allele was sufficient to lighten pigmentation compared to that driven by the *D. americana tan* allele. The molecular mechanism through which noncoding changes in *tan* contribute to pigmentation divergence between *D. americana* and *D. novamexicana* is investigated. *D. melanogaster* transgenics expressing the *D. americana* or *D. novamexicana tan* transgenes were not an amenable system for drawing convincing conclusions regarding mRNA expression level and the impact noncoding sequence changes have on expression. However, *in silico* predictions into changes in transcription factor binding sites

between *D. americana* and *D. novamexicana* revealed three transcription factors with predicted binding sites overlapping a derived sequence change in *D. novamexicana*. These transcription factors and their binding sites represent candidates for a molecular mechanism through which changes in noncoding sequences in these species could contribute to *tan* gene expression and/or phenotypic divergence. I created green fluorescent protein (GFP) reporter genes to test for the presence of enhancer sequences in *D. americana* and *D. novamexicana* intron 1 and intron 3. All constructs GFP expression in *D. melanogaster* transgenics, indicating these noncoding sequences have regulatory ability. Finally, I use intraspecific pigmentation variation within *D. americana* to gain insight into the similarities and differences in phenotypic evolution within and between species. I found that *tan* and *ebony* contribute to pigmentation divergence within *D. americana* for some but not all comparisons, suggesting that additional genes are also involved in the pigmentation variation within *D. americana*. Finally, by comparing phenotypically similar strains of *D. americana*, I uncovered evidence supporting the existence of genetic heterogeneity within *D. americana.* Overall, this research presents important considerations for transgenic analyses using defined genomic integration sites, provides evidence for noncoding DNA sequence in *tan* contributing to phenotypic evolution between *D. americana* and *D. novamexicana*, generates testable hypotheses regarding the molecular mechanism through which noncoding changes in *tan* contribute to pigmentation divergence between *D. americana* and *D. novamexicana*, and offers preliminary data into the genetic loci underlying pigmentation variation within *D. americana*.

Chapter 1

Introduction: Understanding how traits change over time and the contribution of noncoding sequences to phenotypic evolution

Understanding how traits change over time

The amount of biological diversity, or biodiversity, in the world is astonishing. From single-celled bacteria to complex multicellular plants and animals, the number of phenotypes observed is immense. This phenotypic variation is even more impressive when one considers that all of this biodiversity shares a common ancestor. Over time, organisms have changed and adapted leading to the development of novel phenotypes. This includes the often large differences in phenotypes seen between species, as well as the relatively smaller phenotypic variation seen within a single species. At the molecular level, evolution and the development of new phenotypes, at least those that are heritable, are caused by changes in the genetic material. A fundamental goal in biology is understanding how genotypes influence the production of phenotypes seen in the natural world.

Mechanisms of phenotypic evolution

It is understood that differences in phenotypic form, both large and small, have occurred gradually over evolutionary time. Mutations in DNA arise and can be acted upon by adaptive

(i.e. natural selection) and non-adaptive processes (i.e. genetic drift), leading to fixed differences between populations and species. When studying phenotypic evolution, scientists recognize a number of mechanisms at play; these include, but may not be limited to: mutations in coding and noncoding sequences, mutations of large and small effect, gene duplications with subsequent neofunctionalization or subfunctionalization, changes in gene splicing, and influences of genetic linkage and/or pleiotropy. While some general trends have started to emerge regarding these mechanisms, many questions still surround phenotypic evolution (Bush, Chen, Tovar-Corona, & Urrutia, 2017; Piechel & Marquer, 2017).

The role of coding and noncoding sequences in phenotypic evolution

Mutations that could influence phenotypic outputs occur in both coding and noncoding sequences –coding sequences are those that encode the amino acid sequence of proteins in an organism, while noncoding sequences do not encode proteins. These noncoding sequences can play a role in gene expression (controlling when, where, and to what extent a gene is turned on/off), influence splicing (the specific arrangement of coding sequences of a gene), and encode for functional RNA molecules such as transfer RNA, ribosomal RNA, and regulatory RNAs, among other functions. When considering phenotypic evolution, the most well studied functionality of noncoding DNA is in gene expression.

Gene expression can be regulated at the level of transcription or translation. At the transcriptional level of gene regulation, *cis*-regulatory sequences, such as promoters and enhancers, are noncoding DNA sequences that control when, where, and to what extent a gene is turned on/off. These *cis*-regulatory sequences are bound by *trans*-regulatory factors to control mRNA level gene expression. *trans*-regulatory factors are proteins encoded elsewhere in the

genome and often help regulate the expression of more than one gene. In contrast, *cis*-regulatory sequences most often influence the expression of only one gene and often in a highly modular fashion, such that the individual expression patterns of multiple *cis*-regulatory elements lead to the complete expression pattern for the gene they control. As such, a mutation in one *cis*-regulatory sequence may only affect one aspect of a gene's expression – for instance, a change in spatial or temporal patterning. While these changes can still impact phenotypic output of the genes they control, the effect of *cis*-regulatory mutations is hypothesized to be subtler than changes in coding sequences that impact protein function wherever and whenever it is expressed. Relatedly, it is hypothesized that changes in either highly pleiotropic *trans*-regulatory factors and/or coding sequences are more likely to have deleterious effects due to their broad impact. Together, these suggest that *cis*-regulatory sequences are under less evolutionary constraint and can thus evolve faster than coding sequences (Stern & Orgogozo, 2008). Ultimately, these ideas have been synthesized into the "*cis*-regulatory hypothesis" which states that mutations causing morphological variation are expected to arise more often in *cis*-regulatory regions of developmental genes (Stern & Orgogozo, 2008).

Although many evolutionary developmental biologists support the *cis*-regulatory hypothesis (Akam, 1998; Alberch, 1983; Britten & Davidson, 1969, 1971; Carroll, 1995; Jacob, 1977; King & Wilson, 1975; Peter & Davidson, 2011; Wray, 2007; Wray et al., 2003), there has been debate regarding its validity (Alonso & Wilkins, 2005; Hoekstra & Coyne, 2007). The most notable criticism of the *cis*-regulatory hypothesis is that more examples of protein-coding changes impacting phenotypic evolution are known than changes in *cis*-regulatory sequence changes. However, ascertainment bias likely plays a significant role in this trend. More specifically, it is easier to predict the effect of a DNA sequence change in a protein-coding

region due to the known relationship between DNA sequence and protein sequence. In *cis*-regulatory sequences, our ability to predict what impact (if any) a particular mutation has is much weaker. A larger collection of changes in noncoding DNA sequence and their specific effect on phenotypic output will be required in order to improve our ability to predict if/how a non-coding DNA mutation will affect a particular phenotype, as well as lessen the ascertainment bias present in the study of phenotypic evolution. The advent of more powerful experimental approaches to study *cis*-regulatory mutations and precisely edit genomic sequences makes this a more manageable feat, potentially explaining the substantial increase in examples of *cis*-regulatory evolution over the last 20 years (Stern & Orgogozo, 2008).

Despite some debate, the empirical data surrounding phenotypic evolution supports some general trends. When morphological phenotypes are considered, *cis*-regulatory changes are more likely to influence phenotypic evolution; conversely, physiological changes appear to evolve more commonly through changes in coding DNA sequences (Stern & Orgogozo, 2008). Differences in the types of DNA sequence contributing to phenotypic evolution also occurs when looking within and between species; changes in *trans*-regulatory factors and/or coding sequences are more likely to influence phenotypic divergence within a single species, while *cis*-regulatory differences are more common between species (Coolon, McManus, Stevenson, Graveley, & Wittkopp, 2014; Metzger, Wittkopp, & Coolon, 2017; Stern & Orgogozo, 2008; Wittkopp, Haerum, & Clark, 2008). Together these results suggest that phenotypic evolution and its predictability is multifaceted and its study may benefit from more nuanced questions than the relatively simply coding vs non-coding distinction. For example: How does population structure impact phenotypic evolution? How does strength and duration of selection change the type of

mutations fixed within and between populations? When the same phenotype evolves independently, are the same or different genes and/or mutations used?

Pigmentation in *Drosophila* as a model system for studying phenotypic evolution

Pigmentation is an ideal trait for studying phenotypic evolution. Even amongst closely related organisms, pigmentation is one of the most variable traits, providing a wealth of opportunities to study how phenotypic evolution occurs, the genetic basis of this evolution, and its predictability. Studies using pigmentation to study phenotypic evolution exist throughout the tree of life from vertebrates (Greenwood et al., 2011; Hoekstra, 2006; Hoekstra, Hirschmann, Bundey, Insel, & Crossland, 2006; Hoekstra, Krenz, & Nachman, 2005; Nachman, 2005; Parichy, 2006; M. Protas et al., 2008; M. E. Protas et al., 2006; Quigley & Parichy, 2002; Steiner, Weber, & Hoekstra, 2007; Sugie, Terai, Ota, & Okada, 2004), to plants (Cooley, Modliszewski, Rommel, & Willis, 2011; Cooley & Willis, 2009; Holton, 1995; Koes, Verweij, & Quattrocchio, 2005; Martin & Gerats, 1993; Spelt, Quattrocchio, Mol, & Koes, 2002; Winkel-Shirley, 2001), to invertebrates (Bastide et al., 2013; Bickel, Kopp, & Nuzhdin, 2011; Brisson, Templeton, & Duncan, 2004; Dembeck et al., 2015; Endler, Betancourt, Nolte, & Schlotterer, 2016; Hoyal Cuthill & Charleston, 2015; Johnson et al., 2015; Kopp, Graze, Xu, Carroll, & Nuzhdin, 2003; Martinez & Cordeiro, 1970; Nadeau et al., 2016; Pool & Aquadro, 2007; Rebeiz et al., 2009; Salomone, Rogers, Rebeiz, & Williams, 2013; Supple et al., 2013; Wallbank et al., 2016; Wittkopp et al., 2009; Wittkopp, Vaccaro, & Carroll, 2002; Yassin et al., 2016).

Drosophila exhibit a wide pigmentation phenotypes both within and between species (Figure 1.1) and these phenotypes have been studied in the laboratory for over a century. This wealth of research has produced a substantial amount of information regarding the genes

involved in pigment synthesis in *Drosophila* as well as the genes controlling their expression throughout development (Kopp, 2009; Takahashi, 2013; True, 2003; Wittkopp, Carroll, & Kopp, 2003). Combined with the more traditional characteristics making *Drosophila* a model organism for scientific inquiry (short generation time, large progeny size, ease of maintenance in the laboratory), pigmentation in *Drosophila* has become an ideal system to study phenotypic evolution. As a result, there are ever increasing examples of specific genes and mutations contributing to both intraspecies and interspecies pigmentation divergence in multiple *Drosophila* lineages (Massey & Wittkopp, 2016) Furthermore, many of the same pigmentation phenotypes have evolved independently and in different lineages of *Drosophila* (Wittkopp, Williams, Selegue, & Carroll, 2003), providing an ideal system to answer questions regarding the repeatability of phenotypic evolution specifically if the same or different genetic changes are utilized.

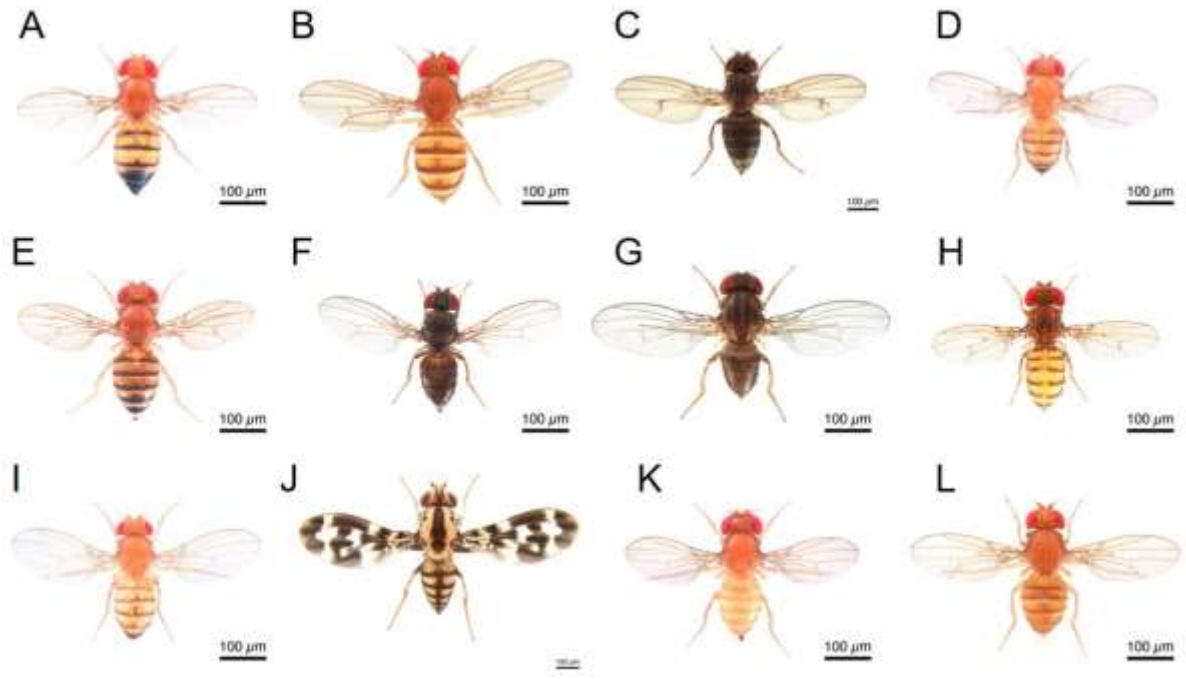

Figure 1.1: Phenotypic diversity in *Drosophila* pigmentation. Representative images of the twelve sequenced species of *Drosophila.* All individuals shown are female. (A) *D. yakuba*, (B) *D. willistoni*, (C) *D. virilis*, (D) *D. simulans*, (E) *D. sechellia*, (F) *D. pseudoobscura*, (G) *D. persimilis*, (H) *D. mojavensis*, (I) *D. melanogaster*, (J) *D. grimshawi*, (K) *D. erecta*, (L). *D. ananassae.* Images taken by Nicolas Gompel and downloaded from flybase.org.

Pigmentation synthesis and regulation in Drosophila

Body color in *Drosophila* is the result of the spatial patterning of four types of pigments: black dopamine melanin, brown dopamine melanin, yellow/tan NBAD scelarotin, and colorless NADA scelarotin (True, 2003; Wittkopp, Carroll, et al., 2003; Wright, 1987). Dietary tyrosine is processed through a biochemical pathway to produce these pigments (Figure 1.2). Tyrosine is converted into DOPA (L-3,4-dihyroxyphenyalanine) by the tyrosine hydroxylase encoded by *pale.* Dopa decarboxylase, an enzyme encoded by *Ddc*, catalyzes a reaction converting DOPA into dopamine. From here, the pathway branches allowing dopamine to be processed into the four pigment types. Dark (black and brown) pigments are produced through processing with phenol oxidases (POs) and the involvement of the protein product of *yellow*. Colorless pigment is created by the conversion of dopamine into N-acetyl dopamine (NADA) via dopamine-acetyl-transferases (DATs) and processing of NADA with POs. Light pigment is produced by the conversion of dopamine into N-β-alanyl dopamine (NBAD) via Ebony and further processing of NBAD with POs. The conversion of dopamine into NBAD is a reversible reaction with Tan facilitating the production of dopamine from NBAD.

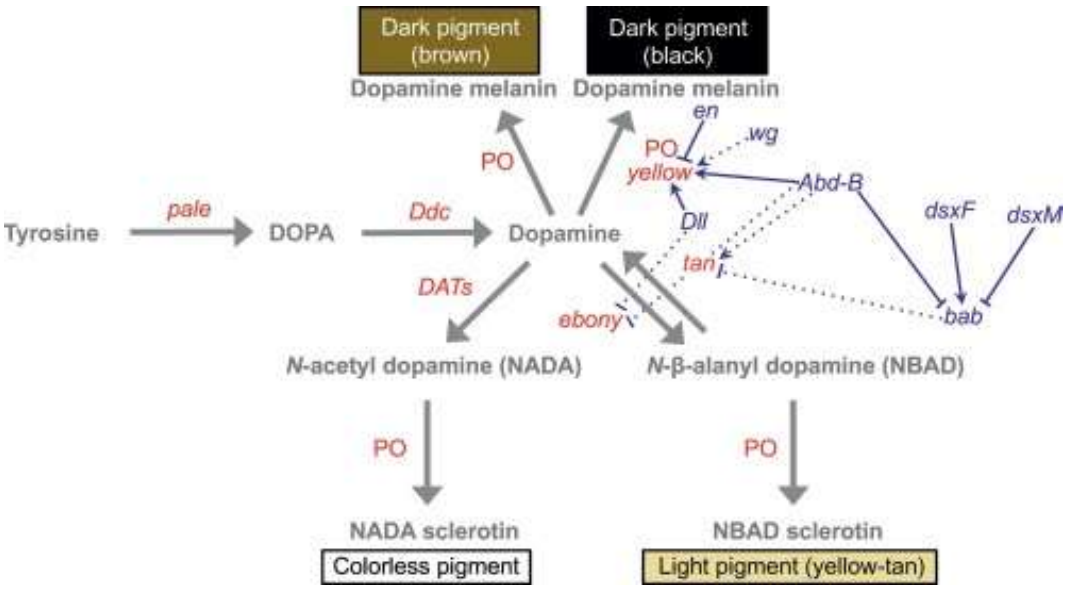

Figure 1.2: Pigmentation biosynthesis in *Drosophila.* From Massey and Wittkopp, 2016. Genes that are part of the pigmentation biosynthesis pathway are shown in red. Genes that are involved in the regulation of pigment development are shown in blue; both direct (solid) and indirect (dashed) regulators are connected with blue lines to the gene(s) they regulate; pointed arrow connections indicate a positive regulatory interaction (i.e. activation) and blunt connection indicate a negative regulatory interaction (i.e. repression). Metabolites are shown in grey. Grey arrows indicate direction of enzymatic reactions in the pathways.

Pigment is deposited into the developing *Drosophila* cuticle in later pupal and early adult stages (Kraminsky et al., 1980; Sugumaran, Giglio, Kundzicz, Saul, & Semensi, 1992; Walter et al., 1996; Wittkopp, Carroll, et al., 2003). The relative expression level and spatial patterning of *yellow*, *tan* and *ebony* work in concert to create the pigmentation patterns seen throughout *Drosophila.* The expression pattern of these pigmentation genes is known to be controlled by at least five transcription factors: *bric-a-brac* (*bab*), *abdominal-B* (*Abd-B*), *doublesex* (*dsx*), *Distal-less* (*Dll*), and *Engrailed* (*en*). These transcription factors have been shown to regulate expression of pigmentation genes both directly and indirectly (Arnoult et al., 2013; Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Jeong, Rokas, & Carroll, 2006; Kopp, Duncan, Godt, & Carroll, 2000; Williams et al., 2008). Direct regulation involves the transcription factor binding to a *cis*-regulatory sequence of a pigmentation gene to control expression, whereas indirect regulation occurs through influence on a gene's direct regulator.

Although specific transcription factors contributing to pigmentation in *Drosophila* have been identified, questions remain surround this pathway and how it contributes to phenotypic evolution. Three such questions are: 1) Do additional transcriptional regulators exist that influence pigmentation? 2) What specific changes in *cis*-regulatory DNA sequences and/or *trans*-regulators of pigmentation genes contribute to phenotypic diversity? 3) How does the genetic basis of pigmentation divergence differ within and between species? Recent RNAi screens suggest that other transcription factors may influence pigmentation in *Drosophila* (Kalay, 2012; Rogers et al., 2014), but we lack information regarding their regulatory targets and examples of these genes contributing to pigmentation divergence in natural populations. Increased numbers of cases in which a particular genetic changes that influence pigmentation and their mechanism of action will increase our understanding of how changes in noncoding DNA sequence and gene regulation generate phenotypic diversity. Lastly, comparing the genetic basis of pigmentation divergence within and between species will allow us better insight into how genetic variation within a species can contribute to divergence between species over evolutionary time. Since inter- and intra-species pigmentation divergence evolve over different timescales, different genes and/or types of genetic changes may underlie differences in pigmentation, even when the pigmentation phenotypes examined are similar (Orr, 2001; Stern & Orgogozo, 2008).

Drosophila americana and *Drosophila novamexicana* as a model for phenotypic evolution

Drosophila americana and *Drosophila novamexicana* represent a tractable system in which to study pigmentation divergence both within and between species. *D. americana* and *D. novamexicana* are interfertile sister species that divergence from a common ancestor ~400,000

years ago (Morales-Hojas, Vieira, & Vieira, 2008). *D. americana* has retained an overall dark pigmentation phenotype common in the virilis species group; *D. novamexicana* has a derived light pigmentation phenotype (Figure 1.3). The pigmentation genes *tan* and *ebony* together explain 87% of the difference in abdominal pigmentation seen between these two species (Wittkopp et al., 2009). Changes in both *cis*- and *trans*-regulatory function controlling the expression of *tan* and *ebony* have been implicated between *D. americana* and *D. novamexicana* (Cooley, Shefner, McLaughlin, Stewart, & Wittkopp, 2012). The specific genetic changes in *tan* and *ebony* responsible for the differences in *cis*-regulatory function have not yet been identified, nor has the molecular mechanism through which these changes act. Additionally, the remaining loci contributing to differences in abdominal pigmentation have not been identified.

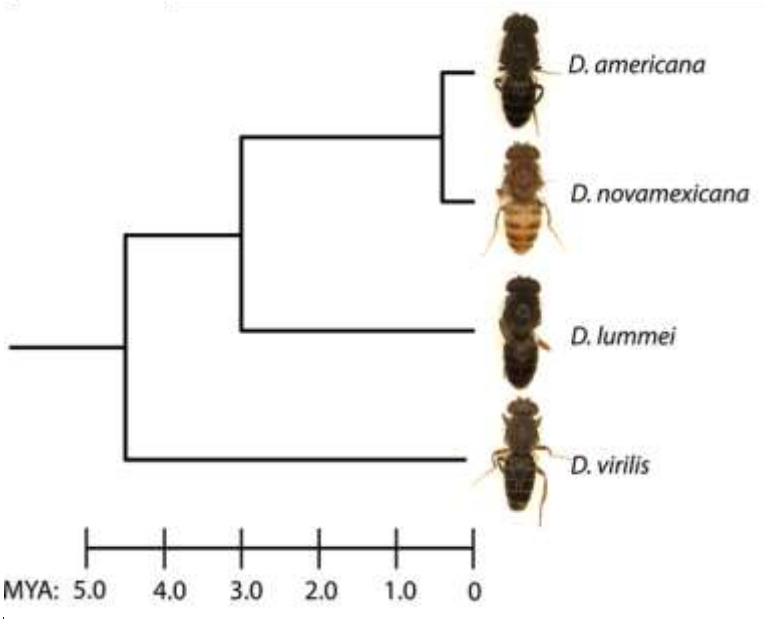

Figure 1.3 Pigmentation in the virilis group of *Drosophila*. Body color for members of the virilis group are shown. *D. americana* has a recently evolved, derived, and dramatically distinct pigmentation phenotype compared to other members of the virilis group. Figure adapted from Cooley et al. 2012.

In addition to the interspecies pigmentation divergence, intraspecies pigmentation diversity exists within *D. americana* with pigmentation varying significantly with longitude

(Wittkopp et al., 2011). As you move west across the continental United States, populations of *D. americana* exhibit lighter pigmentation phenotypes (Figure 1.4). *tan* and *ebony* have been implicated in this intraspecific pigmentation divergence (Wittkopp et al., 2009). As in the between species studies, specific genetic changes in *tan* and *ebony* have not been identified as contributing to pigmentation divergence, nor have additional loci contributing to differences in pigmentation.

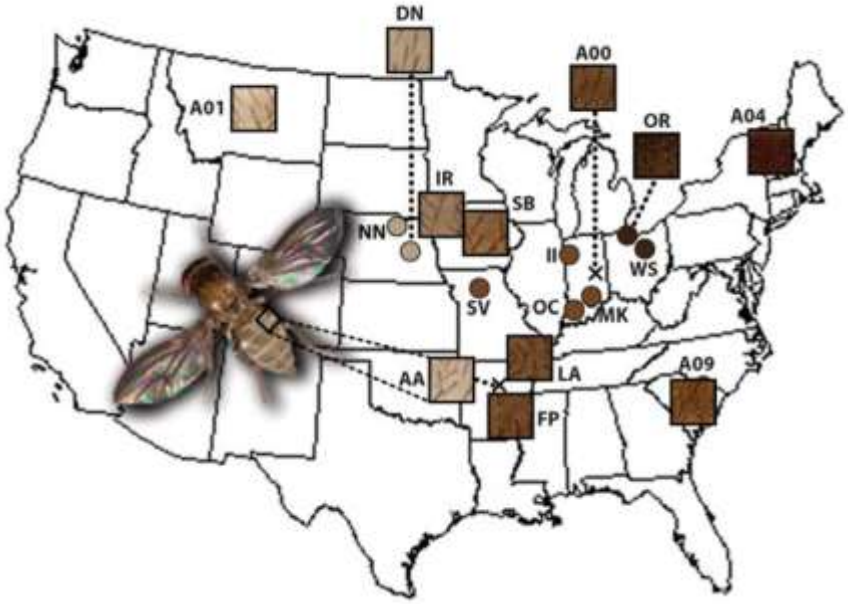

Figure 1.4 Pigmentation variation within *D. americana*. Pigmentation in *D. americana* varies significantly with longitude across the continental United States, with pigmentation becoming lighter moving west. Representative images from the abdominal cuticle of a male fly are shown for each collection site. From Wittkopp et al. 2011.

Overall, the interspecific pigmentation divergence between *D. americana* and *D. novamexicana* combined with the intraspecific pigmentation diversity in *D. americana* provides a unique system in which to study phenotypic evolution. By identifying and comparing the specific genotypes underlying the pigmentation phenotypes in these species, we can gain insight into some of the outstanding questions in the field of phenotypic evolution, including: What

specific genetic changes and molecular mechanisms contribute to pigmentation divergence? And is the genetic basis of pigmentation divergence the same or different between species.

Thesis overview

In the following chapters, the genetic basis of phenotypic evolution is explored using pigmentation divergence between *D. americana* and *D. novamexicana*, as well as within *D. americana*. In Chapter 2, the ability of genomic position to influence the detection of functional differences between alleles is investigated. *D. americana* and *D. novamexicana tan* alleles are compared at multiple genomic locations and the ability to detect pigmentation differences driven by these two alleles is measured. It is shown that genomic positions can influence the ability to detect allelic divergence, which has implications for transgenic analysis of divergent or potentially divergent alleles. In Chapter 3, the role of *tan* in the pigmentation divergence between *D. americana* and *D. novamexicana* is explored using transgenic *tan* alleles. Chimeric alleles of *tan* are created in *D. melanogaster* to investigate the role of previously mapped noncoding sequences in pigmentation divergence. Results suggest that at least one nucleotide that contributes to *tan*'s influence on pigmentation divergence resides in the 5' half of *tan* intron 1. The effect of the chimeric alleles tested appears to be impacted by the functional status of another pigmentation gene *yellow*. In Chapter 4, the molecular mechanism through which noncoding changes in *tan* contribute to pigmentation divergence are explored. mRNA level gene expression of *tan* in the *D. melanogaster* transgenic is measured using pyrosequencing, however, no detectable differences between alleles is found. Previously mapped noncoding sequences (i.e. intron 1 of *tan*) in *D. americana* and *D. novamexicana* are searched for differences in predicted transcription factor binding sites (TFBSs) with some candidate sites identified. Lastly, GFP

reporter genes are constructed for intron 1 and intron 3 of *D. americana* and *D. novamexicana tan* to test these noncoding sequences for the presence of enhancer sequences and for differences in enhancer activity. In Chapter 5, pigmentation divergence is explored within *D. americana*. F2 populations between light and dark lines of *D. americana* are generated and pools of individuals with the most extreme pigmentation phenotypes are genotyped at the candidate loci *tan, ebony,* and *yellow.* Results suggest a role for all three of these genes in intraspecific pigmentation diversity in *D. americana.* The results also suggest genetic heterogeneity exists within *D. americana* such that different genotypes are responsible for the production of similar phenotypes even in strains isolated from the same geographic location.

References

Akam, M. (1998). Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *Int J Dev Biol, 42*(3), 445-451.

Alberch, P. (1983). Development and evolution: embryos, genes, and evolution. *Science, 221*(4607), 257-258. doi:10.1126/science.221.4607.257

Alonso, C. R., & Wilkins, A. S. (2005). The molecular elements that underlie developmental evolution. *Nat Rev Genet, 6*(9), 709-715. doi:10.1038/nrg1676

Arnoult, L., Su, K. F., Manoel, D., Minervino, C., Magrina, J., Gompel, N., & Prud'homme, B. (2013). Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science, 339*(6126), 1423-1426. doi:10.1126/science.1233749

Bastide, H., Betancourt, A., Nolte, V., Tobler, R., Stobe, P., Futschik, A., & Schlotterer, C. (2013). A genome-wide, fine-scale map of natural pigmentation variation in Drosophila melanogaster. *PLoS Genet, 9*(6), e1003534. doi:10.1371/journal.pgen.1003534

Bickel, R. D., Kopp, A., & Nuzhdin, S. V. (2011). Composite effects of polymorphisms near multiple regulatory elements create a major-effect QTL. *PLoS Genet, 7*(1), e1001275. doi:10.1371/journal.pgen.1001275

Brisson, J. A., Templeton, A. R., & Duncan, I. (2004). Population genetics of the developmental gene optomotor-blind (omb) in Drosophila polymorpha: evidence for a role in abdominal pigmentation variation. *Genetics, 168*(4), 1999-2010. doi:10.1534/genetics.104.032078

Britten, R. J., & Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science, 165*(3891), 349-357.

Britten, R. J., & Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol, 46*(2), 111-138.

Bush, S. J., Chen, L., Tovar-Corona, J. M., & Urrutia, A. O. (2017). Alternative splicing and the evolution of phenotypic novelty. *Phil. Trans. R. Soc. B, 372*.

Carroll, S. B. (1995). Homeotic genes and the evolution of arthropods and chordates. *Nature, 376*(6540), 479-485. doi:10.1038/376479a0

Cooley, A. M., Modliszewski, J. L., Rommel, M. L., & Willis, J. H. (2011). Gene duplication in Mimulus underlies parallel floral evolution via independent trans-regulatory changes. *Curr Biol, 21*(8), 700-704. doi:10.1016/j.cub.2011.03.028

Cooley, A. M., Shefner, L., McLaughlin, W. N., Stewart, E. E., & Wittkopp, P. J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in Drosophila americana and D. novamexicana. *Evol Dev, 14*(4), 317-325. doi:10.1111/j.1525-142X.2012.00550.x

Cooley, A. M., & Willis, J. H. (2009). Genetic divergence causes parallel evolution of flower color in Chilean Mimulus. *New Phytol, 183*(3), 729-739. doi:10.1111/j.1469-8137.2009.02858.x

Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R., & Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in Drosophila. *Genome Res, 24*(5), 797-808. doi:10.1101/gr.163014.113

Dembeck, L. M., Huang, W., Magwire, M. M., Lawrence, F., Lyman, R. F., & Mackay, T. F. (2015). Genetic Architecture of Abdominal Pigmentation in Drosophila melanogaster. *PLoS Genet, 11*(5), e1005163. doi:10.1371/journal.pgen.1005163

Endler, L., Betancourt, A. J., Nolte, V., & Schlotterer, C. (2016). Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in Drosophila melanogaster. *Genetics, 202*(2), 843-855. doi:10.1534/genetics.115.183376

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature, 433*(7025), 481-487. doi:10.1038/nature03235

Greenwood, A. K., Jones, F. C., Chan, Y. F., Brady, S. D., Absher, D. M., Grimwood, J., . . . Peichel, C. L. (2011). The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks. *Heredity (Edinb), 107*(2), 155-166. doi:10.1038/hdy.2011.1

Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity (Edinb), 97*(3), 222-234. doi:10.1038/sj.hdy.6800861

Hoekstra, H. E., & Coyne, J. A. (2007). The locus of evolution: evo devo and the genetics of adaptation. *Evolution, 61*(5), 995-1016. doi:10.1111/j.1558-5646.2007.00105.x

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science, 313*(5783), 101-104. doi:10.1126/science.1126121

Hoekstra, H. E., Krenz, J. G., & Nachman, M. W. (2005). Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations. *Heredity (Edinb), 94*(2), 217-228. doi:10.1038/sj.hdy.6800600

Holton, T. A. (1995). Modification of flower colour via manipulation of P450 gene expression in transgenic plants. *Drug Metabol Drug Interact, 12*(3-4), 359-368.

Hoyal Cuthill, J. F., & Charleston, M. (2015). Wing patterning genes and coevolution of Mullerian mimicry in Heliconius butterflies: Support from phylogeography, cophylogeny, and divergence times. *Evolution, 69*(12), 3082-3096. doi:10.1111/evo.12812

Jacob, F. (1977). Evolution and tinkering. *Science, 196*(4295), 1161-1166.

Jeong, S., Rokas, A., & Carroll, S. B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell, 125*(7), 1387-1399. doi:10.1016/j.cell.2006.04.043

Johnson, W. C., Ordway, A. J., Watada, M., Pruitt, J. N., Williams, T. M., & Rebeiz, M. (2015). Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between Drosophila Species. *PLoS Genet, 11*(6), e1005279. doi:10.1371/journal.pgen.1005279

Kalay, G. (2012). *Rapid evolution of cis-regulatory architecture and activity in the Drosophila yellow gene.* (PhD), University of Michigan.

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science, 188*(4184), 107-116.

Koes, R., Verweij, W., & Quattrocchio, F. (2005). Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. *Trends Plant Sci, 10*(5), 236-242. doi:10.1016/j.tplants.2005.03.002

Kopp, A. (2009). Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. *Evolution, 63*(11), 2771-2789. doi:10.1111/j.1558-5646.2009.00761.x

Kopp, A., Duncan, I., Godt, D., & Carroll, S. B. (2000). Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature, 408*(6812), 553-559. doi:10.1038/35046017

Kopp, A., Graze, R. M., Xu, S., Carroll, S. B., & Nuzhdin, S. V. (2003). Quantitative trait loci responsible for variation in sexually dimorphic traits in Drosophila melanogaster. *Genetics, 163*(2), 771-787.

Kraminsky, G. P., Clark, W. C., Estelle, M. A., Gietz, R. D., Sage, B. A., O'Connor, J. D., & Hodgetts, R. B. (1980). Induction of translatable mRNA for dopa decarboxylase in Drosophila: an early response to ecdysterone. *Proc Natl Acad Sci U S A, 77*(7), 4175-4179.

Martin, C., & Gerats, T. (1993). Control of Pigment Biosynthesis Genes during Petal Development. *Plant Cell, 5*(10), 1253-1264. doi:10.1105/tpc.5.10.1253

Martinez, M. N., & Cordeiro, A. R. (1970). Modifiers of Color Pattern Genes in DROSOPHILA POLYMORPHA. *Genetics, 64*(3-4), 573-587.

Massey, J. H., & Wittkopp, P. J. (2016). The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species. *Curr Top Dev Biol, 119*, 27-61. doi:10.1016/bs.ctdb.2016.03.004

Metzger, B. P., Wittkopp, P. J., & Coolon, J. D. (2017). Evolutionary dynamics of regulatory changes underlying gene expression divergence among Saccharomyces species. *Genome Biol Evol*. doi:10.1093/gbe/evx035

Morales-Hojas, R., Vieira, C. P., & 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. *Molecular Ecology, 17*(12), 2910-2926. doi:10.1111/j.1365-294X.2008.03796.x

Nachman, M. W. (2005). The genetic basis of adaptation: lessons from concealing coloration in pocket mice. *Genetica, 123*(1-2), 125-136.

Nadeau, N. J., Pardo-Diaz, C., Whibley, A., Supple, M. A., Saenko, S. V., Wallbank, R. W., . . . Jiggins, C. D. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature, 534*(7605), 106-110. doi:10.1038/nature17961

Orr, H. A. (2001). The genetics of species differences. . *Trends in Ecology & Evolution.*

Parichy, D. M. (2006). Evolution of danio pigment pattern development. *Heredity (Edinb), 97*(3), 200-210. doi:10.1038/sj.hdy.6800867

Peter, I. S., & Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell, 144*(6), 970-985. doi:10.1016/j.cell.2011.02.017

Piechel, C. L., & Marquer, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Phil. Trans. R. Soc. B, 372*.

Pool, J. E., & Aquadro, C. F. (2007). The genetic basis of adaptive pigmentation variation in Drosophila melanogaster. *Molecular Ecology, 16*(14), 2844-2851. doi:10.1111/j.1365-294X.2007.03324.x

Protas, M., Tabansky, I., Conrad, M., Gross, J. B., Vidal, O., Tabin, C. J., & Borowsky, R. (2008). Multi-trait evolution in a cave fish, Astyanax mexicanus. *Evol Dev, 10*(2), 196-209. doi:10.1111/j.1525-142X.2008.00227.x

Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., . . . Tabin, C. J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet, 38*(1), 107-111. doi:10.1038/ng1700

Quigley, I. K., & Parichy, D. M. (2002). Pigment pattern formation in zebrafish: a model for developmental genetics and the evolution of form. *Microsc Res Tech, 58*(6), 442-455. doi:10.1002/jemt.10162

Rebeiz, M., Ramos-Womack, M., Jeong, S., Andolfatto, P., Werner, T., True, J., . . . Carroll, S. B. (2009). Evolution of the tan locus contributed to pigment loss in Drosophila santomea: a response to Matute et al. *Cell, 139*(6), 1189-1196. doi:10.1016/j.cell.2009.11.004

Rogers, W. A., Grover, S., Stringer, S. J., Parks, J., Rebeiz, M., & Williams, T. M. (2014). A survey of the trans-regulatory landscape for Drosophila melanogaster abdominal pigmentation. *Dev Biol, 385*(2), 417-432. doi:10.1016/j.ydbio.2013.11.013

Salomone, J. R., Rogers, W. A., Rebeiz, M., & Williams, T. M. (2013). The evolution of Bab paralog expression and abdominal pigmentation among Sophophora fruit fly species. *Evol Dev, 15*(6), 442-457. doi:10.1111/ede.12053

Spelt, C., Quattrocchio, F., Mol, J., & Koes, R. (2002). ANTHOCYANIN1 of petunia controls pigment synthesis, vacuolar pH, and seed coat development by genetically distinct mechanisms. *Plant Cell, 14*(9), 2121-2135.

Steiner, C. C., Weber, J. N., & Hoekstra, H. E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol, 5*(9), e219. doi:10.1371/journal.pbio.0050219

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution, 62*(9), 2155-2177. doi:10.1111/j.1558-5646.2008.00450.x

Sugie, A., Terai, Y., Ota, R., & Okada, N. (2004). The evolution of genes for pigmentation in African cichlid fishes. *Gene, 343*(2), 337-346. doi:10.1016/j.gene.2004.09.019

Sugumaran, M., Giglio, L., Kundzicz, H., Saul, S., & Semensi, V. (1992). Studies on the enzymes involved in puparial cuticle sclerotization in Drosophila melanogaster. *Arch Insect Biochem Physiol, 19*(4), 271-283. doi:10.1002/arch.940190406

Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., . . . Counterman, B. A. (2013). Genomic architecture of adaptive color pattern divergence and convergence in Heliconius butterflies. *Genome Res, 23*(8), 1248-1257. doi:10.1101/gr.150615.112

Takahashi, A. (2013). Pigmentation and behavior: potential association through pleiotropic genes in Drosophila. *Genes Genet Syst, 88*(3), 165-174.

True, J. (2003). Insect melanism: the molecules matter. *Trends Ecol Evol, 18*(12), 640-647.

Wallbank, R. W., Baxter, S. W., Pardo-Diaz, C., Hanly, J. J., Martin, S. H., Mallet, J., . . . Jiggins, C. D. (2016). Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. *PLoS Biol, 14*(1), e1002353. doi:10.1371/journal.pbio.1002353

Walter, M. F., Zeineh, L. L., Black, B. C., McIvor, W. E., Wright, T. R., & Biessmann, H. (1996). Catecholamine metabolism and in vitro induction of premature cuticle melanization in wild type and pigmentation mutants of Drosophila melanogaster. *Arch Insect Biochem Physiol, 31*(2), 219-233. doi:10.1002/(SICI)1520-6327(1996)31:2<219::AID-ARCH9>3.0.CO;2-U

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell, 134*(4), 610-623. doi:10.1016/j.cell.2008.06.052

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol, 126*(2), 485-493.

Wittkopp, P. J., Carroll, S. B., & Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in Drosophila. *Trends Genet, 19*(9), 495-504. doi:10.1016/S0168-9525(03)00194-X

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2008). Regulatory changes underlying expression differences within and between Drosophila species. *Nat Genet, 40*(3), 346-350. doi:10.1038/ng.77

Wittkopp, P. J., Smith-Winberry, G., Arnold, L. L., Thompson, E. M., Cooley, A. M., Yuan, D. C., . . . McAllister, B. F. (2011). Local adaptation for body color in Drosophila americana. *Heredity (Edinb), 106*(4), 592-602. doi:10.1038/hdy.2010.90

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., . . . Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science, 326*(5952), 540-544. doi:10.1126/science.1176980

Wittkopp, P. J., Vaccaro, K., & Carroll, S. B. (2002). Evolution of yellow gene regulation and pigmentation in Drosophila. *Curr Biol, 12*(18), 1547-1556.

Wittkopp, P. J., Williams, B. L., Selegue, J. E., & Carroll, S. B. (2003). Drosophila pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc Natl Acad Sci U S A, 100*(4), 1808-1813. doi:10.1073/pnas.0336368100

Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet, 8*(3), 206-216. doi:10.1038/nrg2063

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., & Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol, 20*(9), 1377-1419. doi:10.1093/molbev/msg140

Wright, T. R. (1987). The genetics of biogenic amine metabolism, sclerotization, and melanization in Drosophila melanogaster. *Adv Genet, 24*, 127-222.

Yassin, A., Bastide, H., Chung, H., Veuille, M., David, J. R., & Pool, J. E. (2016). Ancient balancing selection at tan underlies female colour dimorphism in Drosophila erecta. *Nat Commun, 7*, 10400. doi:10.1038/ncomms10400

Chapter 2

Sensitivity of allelic divergence to genomic position: Lessons from the *Drosophila tan* gene[1]

Abstract

To identify genetic variants underlying changes in phenotypes within and between species, researchers often utilize transgenic animals to compare the function of alleles in different genetic backgrounds. In *Drosophila*, targeted integration mediated by the ΦC31 integrase allows activity of alternative alleles to be compared at the same genomic location. By using the same insertion site for each transgene, position effects are generally assumed to be controlled for because both alleles are surrounded by the same genomic context. Here, we test this assumption by comparing the activity of *tan* alleles from two *Drosophila* species, *D. americana* and *D. novamexicana*, at five different genomic locations in *D. melanogaster*. We found that the relative effects of these alleles varied among insertion sites, with no difference in activity observed between them at two sites. One of these sites simply silenced both transgenes, but the other allowed expression of both alleles that was sufficient to rescue a mutant phenotype yet failed to reveal the functional differences between the two alleles. These results suggest that more than one insertion site

[1] Research presented in this chapter was published in G3 in July 2016: http://www.g3journal.org/content/early/2016/07/21/g3.116.032029.long

should be used when comparing activity of transgenes because failing to do so could cause functional differences between alleles to go undetected.

Introduction

Understanding the genetic basis of phenotypic change remains a pressing challenge for evolutionary biology. Addressing this challenge requires identifying the genes contributing to phenotypic divergence as well as the specific changes within those genes that alter their function (Stern & Orgogozo, 2008). Linkage mapping and genome-wide association studies (GWAS) are often used to identify regions of the genome associated with phenotypic divergence (Martin & Orgogozo, 2013); however, these approaches must be supplemented with functional tests to demonstrate the phenotypic consequences of individual genes and sequence changes. This functional testing is often accomplished through transgenic analysis that evaluates the effects of a specific gene or region of a gene in different genetic backgrounds.

In *Drosophila*, activity of divergent alleles is typically compared using transgenes inserted into the genome by transposon-mediated transformation (Wittkopp, 2006). Most transposons used for this purpose (e.g., *P-elements*, *piggyBac*, *Hermes*) insert a transgene semi-randomly into the genome (Engels, 1996; Garza, Medhora, Koga, & Hartl, 1991; Guimond, Bideshi, Pinkerton, Atkinson, & O'Brochta, 2003; Handler & Harrell, 1999; Smith, Wohlgemuth, Calvi, Franklin, & Gelbart, 1993; Spradling & Rubin, 1983), which is not ideal because the genomic position of a gene can affect its activity, a phenomenon known as “position effect” (Sturtevant, 1925; Wilson, Bellen, & Gehring, 1990). These position effects can result from chromatin structure at the insertion site (Huisinga et al., 2016; Levis, Hazelrigg, & Rubin, 1985; Wilson et al., 1990) and/or interactions between the sequence of the transgene and surrounding

DNA that affect expression of the transgene (Venken & Bellen, 2007; Wilson et al., 1990). The former generally affects the expression level of the transgene, whereas the latter can impact its expression level and/or spatiotemporal regulation. The extent of position effects has been hypothesized to be the product of two variables: (i) the strength of regulatory elements at the genomic location in which the transgene is inserted, and (ii) the susceptibility of the regulatory sequences in the transgene to altered activity (Wilson et al., 1990). The addition of insulator sequences flanking a transgene can reduce the effects of surrounding genomic context on its activity (Gdula, Gerasimova, & Corces, 1996; Kuhn & Geyer, 2003; Silicheva et al., 2010).

Position effects are especially problematic when comparing activity among transgenes expected to vary in subtle ways. Targeted insertion of transgenes in *Drosophila*, most notably using the bacteriophage ΦC31 integrase system (Groth, Fish, Nusse, & Calos, 2004), can help control for position effects by inserting each transgene of interest into the same genomic position of otherwise identical genomes (Venken & Bellen, 2005). With large collections of "landing sites" (sequences that mediate integration of the transgene) for ΦC31-mediated transformation available (Bateman, Lee, & Wu, 2006; Bischof, Maeda, Hediger, Karch, & Basler, 2007; Venken, He, Hoskins, & Bellen, 2006), this method has become the standard for comparing activity of related alleles in *Drosophila*. Typically, such a study compares a set of transgenic lines in which each transgene is integrated independently into a chosen landing site, with a single landing site used in most cases (Cande, Goltsev, & Levine, 2009; Duncan, Kiefel, & Duncan, 2010; Frankel et al., 2010; Haley, Foys, & Levine, 2010; Joshi, Sun, & Mann, 2010; Kalay & Wittkopp, 2010; Perry, Boettiger, & Levine, 2011; Rebeiz, Jikomes, Kassner, & Carroll, 2011; Sayal, Ryu, & Arnosti, 2011). The use of a single landing site for such studies is justified by the assumption that all alleles compared will be affected similarly by the surrounding genomic

context (Wimmer, 2005). But is this true? Are sets of related transgenes influenced similarly by the surrounding DNA sequence?

Here, we test this assumption by examining the impact of position effects on a comparison of orthologous alleles that contribute to phenotypic divergence between a pair of closely related *Drosophila* species. Specifically, we compare the effects of *tan* alleles from *D. americana* and *D. novamexicana* integrated into the *D. melanogaster* genome at five different genomic locations. *D. americana* and *D. novamexicana* diverged ~400,000 years ago (Caletka & McAllister, 2004; Morales-Hojas, Vieira, & Vieira, 2008) and have evolved dramatic differences in adult pigmentation (Throckmorton, 1982): *D. americana* has a brown body, whereas *D. novamexicana* has a yellow body (Figure 2.1). Prior work has shown that these differences in pigmentation are due in part to divergent sites located in the *tan* gene (Wittkopp et al., 2009). As described below, we found that position effects influenced whether or not a difference in activity could be detected between these two species-specific alleles of *tan*. Further analysis showed that the ability to detect a difference in activity was related to level of expression from the *tan* transgene at each site. These findings suggest that differences between transgenes should be assessed using multiple landing sites.

Figure 2.1: Body color of *D. americana* and *D. novamexicana*. *D. novamexicana* (right) has evolved lighter body pigmentation since it diverged from the common ancestor shared with *D. americana* (left). *D. americana* has retained the darker body pigmentation shared by all other members of the *virilis* group.

Materials and Methods

Generation of Transgenic Flies

Previously constructed transgenes containing *D. americana* or *D. novamexicana tan* (Wittkopp et al., 2009) were injected into *D. melanogaster* using the ΦC31 integrase system. The transgenes contained all exonic and intronic sequences of *tan* as well as 4.1kb of sequence 5' of *tan* and 3.6kb of sequence 3' of *tan* in a *piggyBac* vector (Horn & Wimmer, 2000) containing an attB site used for ΦC31-mediated transformation and Pax6-EGFP, an eye-expressing green fluorescence marker used to detect successful integration (Wittkopp et al., 2009). In addition to non-coding and synonymous changes, these *D. americana* and *D. novamexicana tan* transgenes differ by two amino acids; however, these amino acid differences are not fixed between species and thus unlikely to be responsible for the species-specific differences in pigmentation (Wittkopp et al., 2009). Each *D. melanogaster* host genotype carried a transgene on the X-chromosome using the *vasa cis*-regulatory sequences to express the ΦC31 integrase specifically in the germ-line and a single attP site located on the second (51C – BDSC #24482, 58A – BDSC #24484), third (86Fa – BDSC #24486, 86Fb – BDSC #24749), or fourth (102D – BDSC #24488) chromosome (Bischof et al., 2007). These lines were selected because they contain an eye-expressing red fluorescent protein (RFP) as a visible marker for the landing site; this is in contrast to the majority of strains containing attP landing sites that are commonly used, which use a copy of the *yellow* gene (which restores dark pigmentation in *yellow* mutant flies) as a visible marker for the landing site. GenetiVision (Houston, Texas) performed all DNA preparations and embryo injections according to their standard protocols (http://www.genetivision.com/). Transformant flies (expressing green fluorescent protein in their eyes) were used to establish lines homozygous for each transgene in a *D. melanogaster*

background carrying loss-of-function mutations in the X-linked genes *tan, yellow,* and *white*. The mutant *D. melanogaster tan* allele allowed us to test for rescue of the *tan* mutant pigmentation phenotype by the heterologous *tan* alleles contained in the transgenes; the *yellow* mutant allele reduced the amount of black pigment present in these flies, providing a more sensitive assay for changes in abdominal pigmentation caused by the transgenes; and the *white* allele allowed for easier visualization of the eye-expressing fluorescent transformation marker.

Drosophila husbandry, collection, and abdominal cuticle dissection

For each line to be analyzed, virgin females were mated with males on standard yeast-glucose media at 20°. Upon formation of pupae, parents were removed and the offspring were allowed to continue development. Male offspring were collected 0-1 days post-eclosion and aged to 7-8 days. Flies were stored in 10% glycerol in ethanol prior to dissection.

To harvest abdominal cuticles, 7-8 day old males were removed individually from the 10% glycerol in ethanol solution and placed on a glass slide. Using a razor blade, the abdomen was separated from the rest of the body then cut along the lateral edge parallel with the anterior-posterior axis. The dorsal half of the abdomen was soaked overnight in a solution of phosphate buffered saline (PBS; 1.4 M NaCl, 27 mM KCl, 190 mM Na_2HPO_4, 18 mM KH_2PO_4, adjusted to pH 7.4 with 1 M HCl). After soaking overnight, a single dorsal half of abdominal cuticle was removed from the PBS and placed on a glass slide, dorsal (cuticle) side down. Using forceps, the abdominal cuticle was cleared of any remaining debris. The cleaned cuticle was then mounted dorsal side up in polyvinyl alcohol (PVA) mounting media (BioQuip) on a clean glass slide, covered with a coverslip, and the coverslip was sealed with clear nail polish. This process was repeated for all genotypes analyzed, with 17 – 35 (mean = 27) flies analyzed for each genotype.

To minimize effects of any day-to-day differences in dissections, all genotypes were dissected during each dissection session.

Image collection and processing

Dissected abdominal cuticles were imaged in a single session using a Leica MZ6 microscope and Scion (CFW-1308C) camera operated via TWAIN driver in Adobe Photoshop. Magnification was set to 3.2 with ring light illumination at ~75%. At the beginning of the imaging session, auto white balance (AWB) was used, resulting in a configuration of Gamma 0.605, Red Gain -1,4db, Green Gain 5.4db, and Blue Gain 8.9db with Red Boost and Blue Boost active. These settings were not changed throughout the imaging session. Imaging was conducted at night to minimize changes in ambient lighting. Images were taken slide-by-slide (2 cuticles/slides, cuticles imaged individually) with samples arranged such that no more than two cuticles from the same line were imaged consecutively. A "reference" image of the same dissected cuticle was taken approximately every 10 slides to allow us to evaluate the consistency of the image collection, processing, and analysis pipeline during the multi-hour imaging session.

All images were compiled into a single document in Adobe Photoshop and the "Levels" function was used to adjust the color of all images simultaneously so that the images more closely matched the cuticle appearance visible by eye. This ensured that an identical color adjustment was applied to all photos.

Quantifying pigmentation

Using ImageJ (Rasband, 1997-2016), the area of dorsal abdominal cuticle known as abdominal tergite 4, or "tergite A4", (insert, Figure 2.2) was manually selected using the polygon tool,

excluding any regions containing cracks, holes, or overlapping regions. Measurements of pixel intensity (area, mean, standard deviation, mode, min, max, and median) were taken for each selection. These results were compiled into a single Microsoft Excel spreadsheet where other identifying information was then added, including imaging order (ranging from 1 to 479), allele (no transgene control, *D. americana*, *D. novamexicana*), and landing site (control, 51C, 58A, 86Fa, 86Fb, 102D). Since ImageJ quantifies pigmentation (pixel intensity of a grayscale image) on a 0-255 scale (dark-light), we subtracted the reported pixel intensity from 255 so that darker cuticles had a higher pigmentation score. This file was then saved as a .csv file for statistical analysis in R.

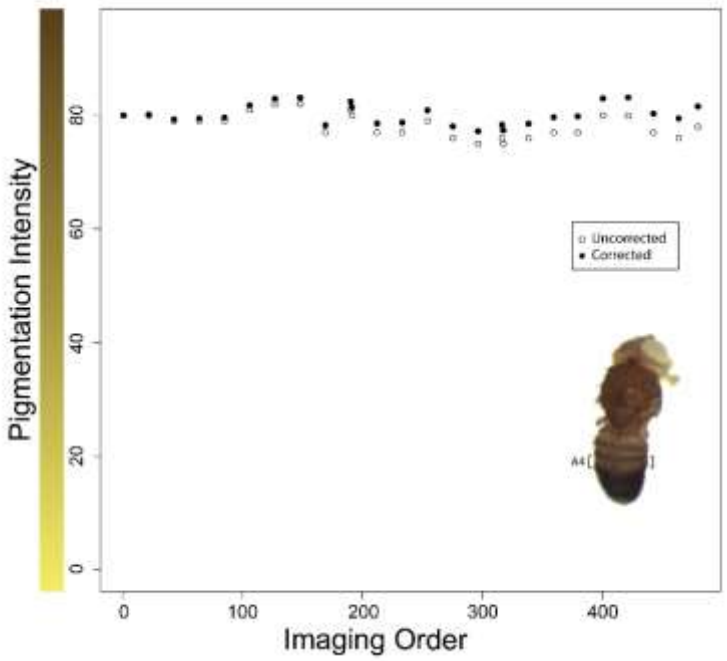

Figure 2.2: Measurements of pigmentation intensity in a control sample varied slightly during image collection. Raw median pigmentation intensity in tergite A4 (insert) is plotted against imaging order for the reference cuticle (open circles). All images were taken during in a single sitting without adjustment of lighting, focus, or other imaging parameters; the small ($\beta = -0.0075$), yet significant (p-value = 0.008), downward trend in pigmentation intensity as the imaging progressed, presumably as a result of changes in ambient lighting or other uncontrolled imaging variables. An imaging order correction was therefore applied to all measurements, as described in the *Materials and Methods* section. Corrected median pigmentation intensity values for the same images are also plotted against imaging order (closed circles) to show the effects of this correction.

Data analysis

Median pigmentation intensity of tergite A4 for each sample reported by ImageJ was analyzed using R v3.2.5 (RCoreTeam, 2016). Median pigmentation was chosen for analysis instead of mean pigmentation intensity to minimize the impact of outlier (excessively white or black) pixels.

To test for systematic changes in imaging conditions that might have occurred during the imaging session, a linear regression was performed comparing median pigmentation values from the reference cuticle and the image order number. A small but significant regression coefficient (beta = -0.0075, p-value = 0.008) was observed, so a correction for imaging order was applied to each median by subtracting (image order number * -0.0075) from the original median value. The differences in reference cuticle values before and after applying this correction are shown in Figure 2.2. Note that all analyses described below were also performed on data without this correction and produced the same pattern of statistically significant results (data not shown).

Median pigmentation intensity of tergite A4 for each sample reported by ImageJ was then fitted to the following model to test effects of landing site, allele, and the interaction between the two:

$$Y_{ijk} = site_i + allele_j + site{\times}allele_{ij} + e_{ijk}$$

Pairwise t-tests using unpooled standard deviations were then performed on the corrected pigmentation medians to identify which comparisons among *tan* alleles and/or landing sites were statistically significant. Statistical significance was assessed using p-values adjusted for multiple

testing by the Benjamini and Hochberg method (Benjamini & Hochberg, 1995) as implemented in the pairwise.t.test function in R.

Measuring relative expression of tan *transgenes at each genomic location*

To test for differences in expression level of transgenes inserted at each genomic location in *D. melanogaster*, relative expression levels of the *D. americana tan* transgene were measured using pyrosequencing (Wittkopp, 2011). Specifically, we measured the mRNA abundance of the *D. americana tan* allele inserted at each genomic location relative to the mRNA abundance of the *D. novamexicana tan* allele inserted at the 86Fa landing site. This *D. novamexicana* genotype was chosen as the internal reference point for measurements of *D. americana tan* expression because it caused an intermediate pigmentation phenotype, suggesting it might also have an intermediate level of expression. The P14-P15 pupal stage was analyzed because *D. americana* and *D. novamexicana tan* have previously been shown to be most highly expressed during this time (Cooley, Shefner, McLaughlin, Stewart, & Wittkopp, 2012). Pupal heads and wings were removed to avoid measuring *tan* expression in those tissues, focusing our measurements on expression in the thorax and abdomen where pigmentation phenotypes are most apparent.

For each landing site, both genomic DNA and total RNA were extracted from three replicate samples, each containing six dissected pupae expressing *D. americana tan* and six dissected pupae expressing *D. novamexicana tan*. cDNA was reverse transcribed from extracted RNA using a polyT primer for each sample. Both genomic DNA and cDNA were analyzed by pyrosequencing as described in Wittkopp et al. (Wittkopp, 2011). PCR primers used to amplify the sequence analyzed (which was located in exon 7) were 5'-GATGCTGAAGTCCAGCGTGTC-3' and 5'-biotin- CAGCCGCCAGTGACATCA-3', and the

primer used for pyrosequencing had the sequence 5'- CGAGCACGATGTCCG-3'. All measurements were then normalized to the relative expression of the *D. americana tan* transgene inserted at landing site 86Fa to compare expression among the D. *americana tan* transgenes at different landing sites.

Results and Discussion

To test the assumption that position effects are negligible when comparing divergent alleles of the same gene at a single genomic location, we transformed *D. americana* and *D. novamexicana tan* alleles into five different genomic locations in *D. melanogaster* (51C, 58A, 86Fa, 86Fb, and 102D). Each of these transgenic lines was then crossed with *D. melanogaster yellow, white, tan* mutants (see methods for full genotype) to move the transgenes into genetic backgrounds lacking a functional copy of the *D. melanogaster tan* gene. Prior work has shown that the difference in body color seen between *D. americana* and *D. novamexicana* (Figure 2.1) is due in part to changes in *tan* and that these *D. americana* and *D. novamexicana tan* transgenes significantly increase abdominal pigmentation in a *D. melanogaster tan* mutant (Wittkopp et al., 2009). The transgenic *tan* allele from the more darkly pigmented *D. americana* was reported to increase pigmentation significantly more than the transgenic *tan* allele from the more lightly pigmented *D. novamexicana*, indicating that there is functional divergence between these species-specific alleles that affects pigmentation (Wittkopp et al., 2009).

To determine whether the insertion site of the *D. americana* and *D. novamexicana tan* transgenes affected their relative activity, we used an analysis of variance (ANOVA) to test for significant effects on pigmentation of allelic identity of the *tan* transgene (*D. americana* or *D. novamexicana*), genomic location of the landing site, and the interaction between the two. All

three factors were found to be statistically significant predictors of pigmentation intensity (Table 2.1). In other words, pigmentation differences were detected between alleles and among landing sites, and the difference between alleles differed among landing sites. The significance of this interaction term is particularly interesting because it suggests that the effects of genomic context might differ between alleles, implying that the landing site used to compare the function of *D. americana* and *D. novamexicana* alleles might alter the conclusions drawn about differences (or lack thereof) between these two alleles.

Table 2.1 Analysis of variance (ANOVA) in pigmentation indicates that *tan* allelic identity, genomic location, and the interaction between allele and genomic location affect pigmentation intensity.

Source of variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F value (F_0)**	**P-value**
1) *tan* transgene identity	2	3.08×10^4	1.54×10^4	1.54×10^2	2.76×10^{-46}
2) Landing site	4	5.74×10^4	1.44×10^4	1.43×10^2	1.89×10^{-67}
3) Interaction between 1 & 2	4	2.54×10^3	6.34×10^2	6.33	6.73×10^{-5}
4) Residuals	290	2.90×10^4	90.5	N/A	N/A

One way that the genomic context can affect a transgene is to simply silence it. To determine whether such silencing was contributing to the difference in allelic differences observed among insertion sites, we used t-tests to determine whether each transgene caused a statistically significant darkening of pigmentation in each transgenic line relative to the *D. melanogaster tan* mutant phenotype. Such a darkening would indicate that the transgene carried was being expressed at a level sufficient to restore at least some dark pigmentation in *D. melanogaster*. We found that the transgenic *tan* alleles from both *D. americana* and *D. novamexicana* failed to significantly alter pigmentation of the *D. melanogaster tan* mutant when inserted into the fourth

chromosome at cytological position 102D (Table 2.2). This evidence of transgene silencing is consistent with prior studies showing that the fourth chromosome of *D. melanogaster* is highly heterochromatic (Riddle & Elgin, 2006; Riddle, Shaffer, & Elgin, 2009) and can suppress expression of transgenes (Salzler et al., 2013; Sun et al., 2000). Landing site 102D does not always silence transgenes, however; other transgenes inserted into the 102D landing site have been shown to be expressed during larval stages (Bischof et al., 2007; Barolo & Evans, personal communication). At each of the other four landing sites tested (all located on chromosome 2 or chromosome 3), both the *D. americana* and *D. novamexicana tan* transgenes caused a significant darkening of pigmentation relative to the *tan* mutant phenotype (Table 2.2), indicating that the transgenes were expressed and producing functional Tan protein. To determine whether the silencing of transgenes at landing site 102D was sufficient to explain the significant interaction observed between transgene identity and landing site in the initial ANOVA, we excluded flies with transgenes inserted into this site and repeated this ANOVA. We found that the two main effect terms (transgene identity and landing site) and the interaction term remained statistically significant (Table 2.3), indicating that the relative activity of the *D. americana* and *D. novamexicana tan* transgenes differed even among sites that allowed expression of both transgenes.

Table 2.2: **Table 2. Pairwise t-tests show which transgenes inserted at which insertion sites alter pigmentation relative to D. melanogaster tan mutants as well as which landing sites show evidence of functional differences between the D. americana and D. novamexicana tan alleles.** P-values adjusted by the Benjamini and Hochberg method from all possible pairwise t-tests using unpooled standard deviation are shown for each pair of genotypes compared. The *tan* mutant column shows results from comparisons between each transgenic genotype and the *tan* mutant (no transgene) control. Note that neither transgene darkened pigmentation relative to the *tan* mutant when inserted at landing site 102D. Shaded boxes indicate comparisons between the *D. americana* and *D. novamexicana tan* alleles inserted at the same landing site. Significant differences ($p<0.05$) in median pigmentation were observed for transgenes inserted at 58A, 86Fa, and 86Fb, but not 51C or 102D. *D. amer* = *D. americana*, *D. nova* = *D. novamexicana*.

		tan mutant	51C *D. amer*	51C *D. nova*	58A *D. amer*	58A *D. nova*	86Fa *D. amer*	86Fa *D. nova*	86Fb *D. amer*	86Fb *D. nova*	102D *D. amer*
51C	*D. amer*	$4x10^{-9}$	-	-	-	-	-	-	-	-	-
51C	*D. nova*	$8x10^{-8}$	0.411	-	-	-	-	-	-	-	-
58A	*D. amer*	$5x10^{-18}$	$3x10^{-10}$	$2x10^{-11}$	-	-	-	-	-	-	-
58A	*D. nova*	$4x10^{-12}$	$1x10^{-3}$	$1x10^{-4}$	$5x10^{-5}$	-	-	-	-	-	-
86Fa	*D. amer*	$4x10^{-24}$	$3x10^{-14}$	$1x10^{-15}$	0.317	$2x10^{-7}$	-	-	-	-	-
86Fa	*D. nova*	$2x10^{-21}$	$4x10^{-7}$	$1x10^{-8}$	$1x10^{-3}$	0.093	$2x10^{-6}$	-	-	-	-
86Fb	*D. amer*	$3x10^{-23}$	$1x10^{-15}$	$9x10^{-17}$	0.012	$2x10^{-9}$	0.064	1×10^{-8}	-	-	-
86Fb	*D. nova*	$9x10^{-17}$	$1x10^{-7}$	$7x10^{-9}$	0.064	0.011	$2x10^{-3}$	0.195	$2x10^{-5}$	-	-
102D	*D. amer*	0.125	$1x10^{-4}$	$1x10^{-3}$	$9x10^{-14}$	$3x10^{-8}$	$4x10^{-15}$	$8x10^{-11}$	$9x10^{-17}$	$1x10^{-11}$	-
102D	*D. nova*	0.088	$3x10^{-7}$	$7x10^{-6}$	$7x10^{-17}$	$8x10^{-11}$	$2x10^{-23}$	$3x10^{-20}$	$3x10^{-22}$	$2x10^{-15}$	0.706

Table 2.3: Table 3. Interaction between allelic identity and genomic location remains significant after excluding silenced transgenes. Results from ANOVA after excluding flies with transgenes inserted at landing site 102D are shown.

Source of variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F value (F_0)**	**P-value**
1) *tan* transgene identity	2	3.53×10^4	1.77×10^4	1.66×10^2	6.39×10^{-46}
2) Landing site	3	2.41×10^4	8.05×10^3	75.5	2.00×10^{-34}
3) Interaction between 1 & 2	3	1.54×10^3	5.13×10^2	4.82	2.83×10^{-3}
4) Residuals	240	2.56×10^4	1.07×10^2	N/A	N/A

To further investigate this difference in relative transgene activity among insertion sites, we used a series of t-tests to compare the pigmentation phenotype caused by the *D. americana* and *D. novamexicana tan* alleles inserted at the same landing site. We found that the *D. americana tan* allele increased dark pigmentation of the *D. melanogaster tan* mutant significantly more than the *D. novamexicana tan* allele when inserted at three (58A, 86Fa and 86Fb) of the four landing sites expressing the transgenes (Table 2.2, Figure 2.3). The difference in activity between these two alleles was masked, however, when then transgenes were inserted into the landing site at 51C ($p = 0.411$, Table 2.2, Figure 2.3). Excluding flies with transgenes at this landing site (51C) as well as flies with transgenes at the landing site that silenced the transgenes (102D) from the ANOVA described above resulted in a non-significant interaction between transgene allele and insertion site (Table 2.4), indicating that the relative effects of the *D. americana* and *D. novamexicana tan* transgenes on pigmentation were comparable at the 58A, 86Fa, and 86Fb landing sites.

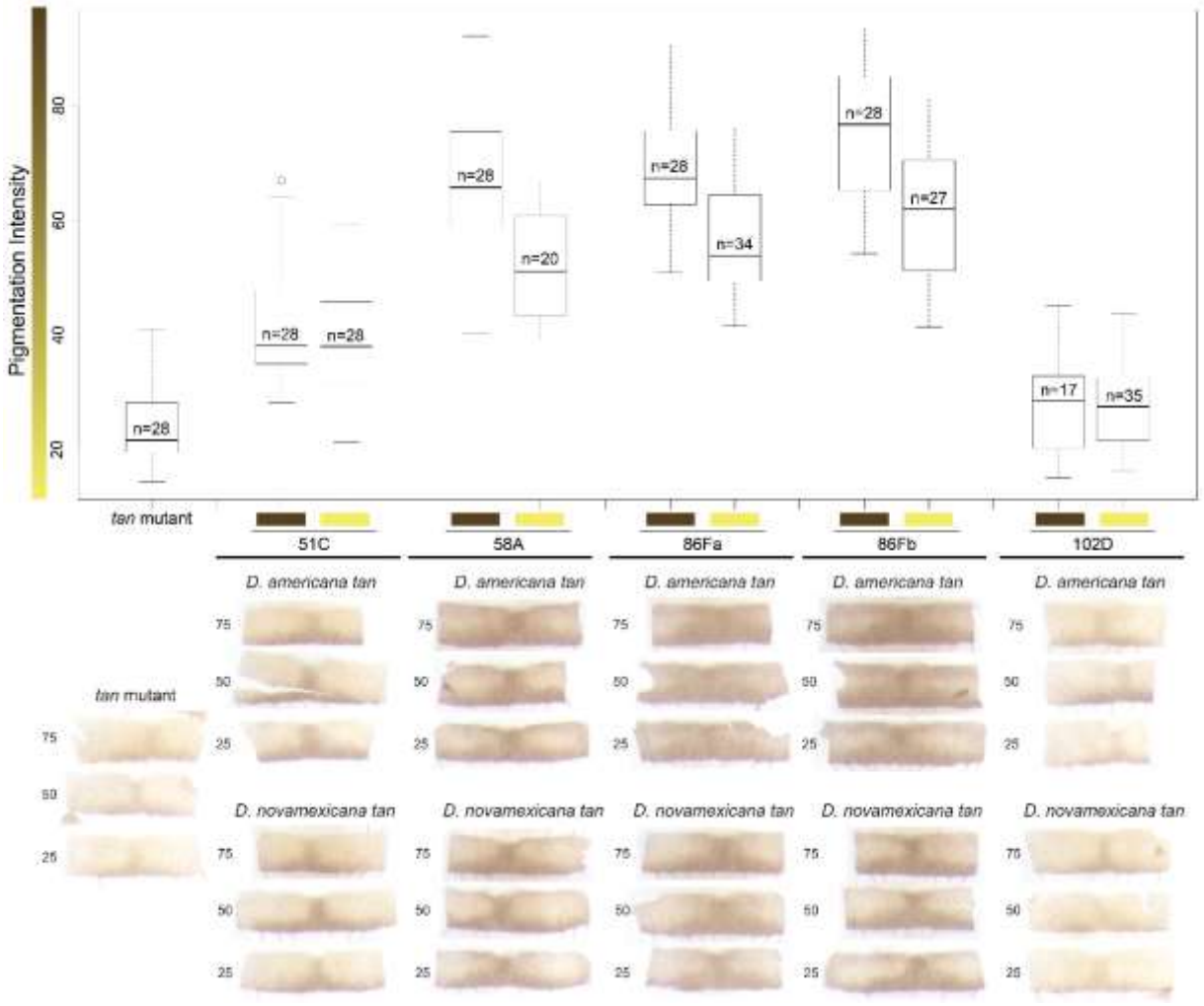

Figure 2.3 Genomic location can impact the relative difference in pigmentation caused by *D. americana* and *D. novamexicana tan* alleles. Box plots summarizing the pigmentation phenotypes observed for each genotype. The median (center line), first quartile (bottom of box), third quartile (top of box), and +/- 1.5x the interquartile range (whiskers) are shown for each genotype examined. Yellow boxes along the x-axis represent the *D. novamexicana* allele and brown boxes represent the *D. americana* allele. Significant increases in pigmentation from the control were detected for all genomic locations except 102D (Table 2.2). Three of the other four landing sites (58A, 86Fa, 86Fb) showed significant differences in pigmentation driven by the *D. americana* and *D. novamexiana tan* alleles whereas the fourth landing site (51C) did not show a detectable difference in pigmentation between flies carrying the two species' alleles (Table 2.2). Representative images from the 25th percentile (first quartile), median, and 75thpercentile (third quartile) are shown below the box plot for each genotype. The most striking differences between alleles are seen in the anterior regions outside the dorsal midline stripe.

Prior work has shown that position effects often alter expression levels of transgenes (Markstein, Pitsouli, Villalta, Celniker, & Perrimon, 2008; Namciu, Blochlinger, & Fournier, 1998; Wilson et al., 1990), thus we hypothesized that the different pigmentation phenotypes

resulting from different insertion sites of the transgenes might be caused by differences in transgene expression among landing sites. To test this hypothesis, we used pyrosequencing to measure the relative expression of the *D. americana tan* transgene among landing sites (Figure 2.4). Genomic locations (58A, 86Fa, 86Fb) that showed statistically significant differences in pigmentation caused by the *D. americana* and *D. novamexicana tan* alleles had the highest levels of *D. americana tan* expression. The genomic location (51C) in which the *D. americana* and *D. novamexicana tan* alleles showed a significant increase in pigmentation relative to the *D. melanogaster tan* mutant, but no differences in pigmentation between flies carrying the *D. americana* and *D. novamexicana tan* alleles had a lower level of *D. americana tan* expression. The genomic location (102D) in which neither the *D. americana* nor the *D. novamexicana tan* transgene increased pigmentation significantly relative to the *D. melanogaster tan* mutant showed the lowest expression of *D. americana tan* among all five lines. These results confirm that different landing sites resulted in different levels of transgene expression and suggest that a threshold in transgene expression level must be reached before the different activities of the *D. americana* and *D. novamexicana tan* transgenes can be detected. We expect that this will be generally true when comparing activities of divergent alleles inserted into the same genomic location, but that the value of this threshold will likely differ depending on the strength of regulatory sequences in the transgenes, genomic context, and/or the relative difference in activity between alleles.

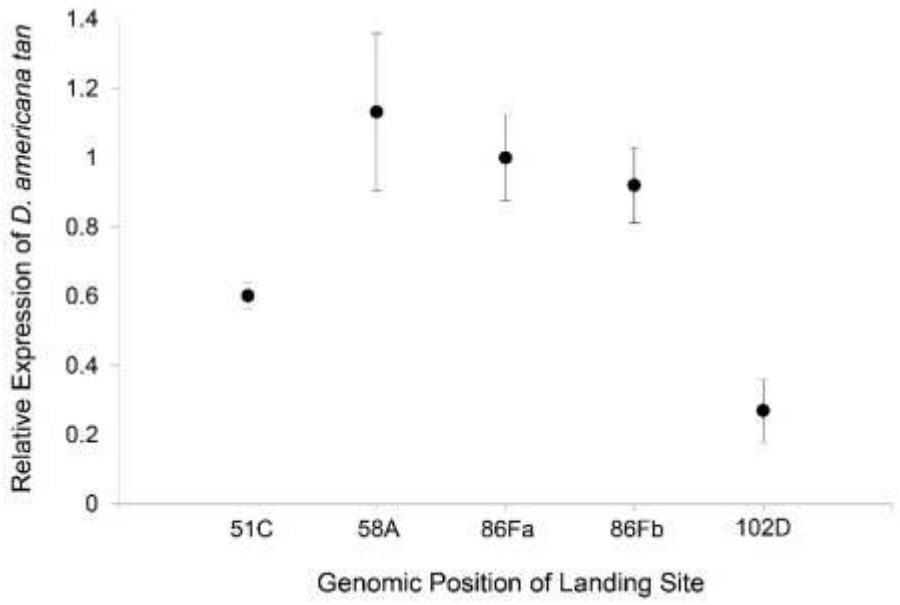

Figure 2.4: Genomic location impacts relative expression of the *D. americana tan* transgene in *D. melanogaster*. Expression of the *D. americana tan* transgene inserted at each of the five landing sites tested is shown relative to its expression when inserted in the 86Fa landing site. Circles indicate mean expression among replicate samples and the error bars show the 95% C.I. of the estimates. Note that the relative expression level of *D. americana tan* among landing sites correlates with the ability to detect differences in abdominal pigmentation (Figure 3). The *D. americana tan* transgene inserted at 58A, 86Fa, and 86Fb all showed similar expression as well as similar pigmentation phenotypes. The D. *americana tan* transgene inserted at 51C had a level of expression between these lines and the line with the transgene inserted at 102D, as well as pigmentation that was intermediate between these lines and 102D. The *D. americana tan* transgene inserted at 102D had the lowest transgene expression and failed to increase dark pigmentation relative to the *tan* mutant phenotype.

In summary, by comparing activities of divergent alleles of the same gene at five different genetic locations, we were able to test the assumption that position effects can be ignored as long as the two alleles compared are inserted into the same genomic location and the transgenes are expressed. We found this not to be true; *D americana* and *D. novamexicana tan* transgenes inserted at landing site (51C) increased dark pigmentation relative to a *D. melanogaster tan* mutant, yet showed no significant difference in their relative activity. If we had only compared the effects of these *tan* alleles at the 51C landing site, we would have concluded that they had conserved functions. The lower level of transgene expression at this site relative to transgenes inserted at the three landing sites that allowed a functional difference between the *D. americana* and *D. novamexicana tan* alleles to be detected suggests that landing sites allowing

the highest levels of transgene expression might provide the most power for detecting differences between alleles. We recommend that at least three genomic locations should be tested to search for allelic differences in activity. Although this increased production of transgenic lines would increase cost and workload, they would help prevent inaccurate conclusions from being drawn from transgenes affected by position effects.

Acknowledgments

We would like to thank Scott Barolo and members of the Wittkopp lab for helpful discussions as well as Robert Dikeman for experimental assistance. This work was funded by the National Institute of Health (R01-GM-089736).

References

Bateman, J. R., Lee, A. M., & Wu, C. T. (2006). Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. *Genetics, 173*(2), 769-777. doi:10.1534/genetics.106.056945

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289-300.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., & Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A, 104*(9), 3312-3317. doi:10.1073/pnas.0611511104

Caletka, B. C., & McAllister, B. F. (2004). A genealogical view of chromosomal evolution and species delimitation in the Drosophila virilis species subgroup. *Molecular Phylogenetics and Evolution, 33*(3), 664-670. doi:10.1016/j.ympev.2004.08.007

Cande, J., Goltsev, Y., & Levine, M. S. (2009). Conservation of enhancer location in divergent insects. *Proc Natl Acad Sci U S A, 106*(34), 14414-14419. doi:10.1073/pnas.0905754106

Cooley, A. M., Shefner, L., McLaughlin, W. N., Stewart, E. E., & Wittkopp, P. J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in Drosophila americana and D. novamexicana. *Evol Dev, 14*(4), 317-325. doi:10.1111/j.1525-142X.2012.00550.x

Duncan, D., Kiefel, P., & Duncan, I. (2010). Control of the spineless antennal enhancer: direct repression of antennal target genes by Antennapedia. *Dev Biol, 347*(1), 82-91. doi:10.1016/j.ydbio.2010.08.012

Engels, W. R. (1996). P elements in Drosophila. *Curr Top Microbiol Immunol, 204*, 103-123.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature, 466*(7305), 490-493. doi:10.1038/nature09158

Garza, D., Medhora, M., Koga, A., & Hartl, D. L. (1991). Introduction of the transposable element mariner into the germline of Drosophila melanogaster. *Genetics, 128*(2), 303-310.

Gdula, D. A., Gerasimova, T. I., & Corces, V. G. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proc Natl Acad Sci U S A, 93*(18), 9378-9383.

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics, 166*(4), 1775-1782.

Guimond, N., Bideshi, D. K., Pinkerton, A. C., Atkinson, P. W., & O'Brochta, D. A. (2003). Patterns of Hermes transposition in Drosophila melanogaster. *Mol Genet Genomics, 268*(6), 779-790. doi:10.1007/s00438-002-0800-4

Haley, B., Foys, B., & Levine, M. (2010). Vectors and parameters that enhance the efficacy of RNAi-mediated gene disruption in transgenic Drosophila. *Proc Natl Acad Sci U S A, 107*(25), 11435-11440. doi:10.1073/pnas.1006689107

Handler, A. M., & Harrell, R. A., 2nd. (1999). Germline transformation of Drosophila melanogaster with the piggyBac transposon vector. *Insect Mol Biol, 8*(4), 449-457.

Horn, C., & Wimmer, E. A. (2000). A versatile vector set for animal transgenesis. *Dev Genes Evol, 210*(12), 630-637.

Huisinga, K. L., Riddle, N. C., Leung, W., Shimonovich, S., McDaniel, S., Figueroa-Clarevega, A., & Elgin, S. C. (2016). Targeting of P-Element Reporters to Heterochromatic Domains by Transposable Element 1360 in Drosophila melanogaster. *Genetics, 202*(2), 565-582. doi:10.1534/genetics.115.183228

Joshi, R., Sun, L., & Mann, R. (2010). Dissecting the functional specificities of two Hox proteins. *Genes Dev, 24*(14), 1533-1545. doi:10.1101/gad.1936910

Kalay, G., & Wittkopp, P. J. (2010). Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among Drosophila species. *PLoS Genet, 6*(11), e1001222. doi:10.1371/journal.pgen.1001222

Kuhn, E. J., & Geyer, P. K. (2003). Genomic insulators: connecting properties to mechanism. *Curr Opin Cell Biol, 15*(3), 259-265.

Levis, R., Hazelrigg, T., & Rubin, G. M. (1985). Effects of genomic position on the expression of transduced copies of the white gene of Drosophila. *Science, 229*(4713), 558-561.

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S. E., & Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. *Nat Genet, 40*(4), 476-483. doi:10.1038/ng.101

Martin, A., & Orgogozo, V. (2013). The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution, 67*(5), 1235-1250. doi:10.1111/evo.12081

Morales-Hojas, R., Vieira, C. P., & 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. *Molecular Ecology, 17*(12), 2910-2926. doi:10.1111/j.1365-294X.2008.03796.x

Namciu, S. J., Blochlinger, K. B., & Fournier, R. E. (1998). Human matrix attachment regions insulate transgene expression from chromosomal position effects in Drosophila melanogaster. *Mol Cell Biol, 18*(4), 2382-2391.

Perry, M. W., Boettiger, A. N., & Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci U S A, 108*(33), 13570-13575. doi:10.1073/pnas.1109873108

Rasband, W. S. (1997-2016). ImageJ. Bethesda, Maryland, USA: U. S. National Institutes of Health. Retrieved from http://imagej.nih.gov/ij/

RCoreTeam. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rebeiz, M., Jikomes, N., Kassner, V. A., & Carroll, S. B. (2011). Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A, 108*(25), 10036-10043. doi:10.1073/pnas.1105937108

Riddle, N. C., & Elgin, S. C. R. (2006). The dot chromosome of Drosophila: Insights into chromatin states and their change over evolutionary time. *Chromosome Research, 14*(4), 405-416. doi:10.1007/s10577-006-1061-6

Riddle, N. C., Shaffer, C. D., & Elgin, S. C. R. (2009). A lot about a little dot - lessons learned from Drosophila melanogaster chromosome 4. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire, 87*(1), 229-241. doi:10.1139/O08-119

Salzler, H. R., Tatomer, D. C., Malek, P. Y., McDaniel, S. L., Orlando, A. N., Marzluff, W. F., & Duronio, R. J. (2013). A sequence in the Drosophila H3-H4 Promoter triggers histone locus body assembly and biosynthesis of replication-coupled histone mRNAs. *Dev Cell, 24*(6), 623-634. doi:10.1016/j.devcel.2013.02.014

Sayal, R., Ryu, S. M., & Arnosti, D. N. (2011). Optimization of reporter gene architecture for quantitative measurements of gene expression in the Drosophila embryo. *Fly (Austin), 5*(1), 47-52.

Silicheva, M., Golovnin, A., Pomerantseva, E., Parshikov, A., Georgiev, P., & Maksimenko, O. (2010). Drosophila mini-white model system: new insights into positive position effects and the role of transcriptional terminators and gypsy insulator in transgene shielding. *Nucleic Acids Res, 38*(1), 39-47. doi:10.1093/nar/gkp877

Smith, D., Wohlgemuth, J., Calvi, B. R., Franklin, I., & Gelbart, W. M. (1993). hobo enhancer trapping mutagenesis in Drosophila reveals an insertion specificity different from P elements. *Genetics, 135*(4), 1063-1076.

Spradling, A. C., & Rubin, G. M. (1983). The effect of chromosomal position on the expression of the Drosophila xanthine dehydrogenase gene. *Cell, 34*(1), 47-57.

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution, 62*(9), 2155-2177. doi:10.1111/j.1558-5646.2008.00450.x

Sturtevant, A. H. (1925). The Effects of Unequal Crossing over at the Bar Locus in Drosophila. *Genetics, 10*(2), 117-147.

Sun, F. L., Cuaycong, M. H., Craig, C. A., Wallrath, L. L., Locke, J., & Eigin, S. C. R. (2000). The fourth chromosome of Drosophila melanogaster: Interspersed euchromatic and heterochromatic domains. *Proceedings of the National Academy of Sciences of the United States of America, 97*(10), 5340-5345. doi:DOI 10.1073/pnas.090530797

Throckmorton, L. H. (1982). In C. Ashburner, M. H. L., Thompson, J. N. (Ed.), *The Genetics and Biology of Drosophila* (Vol. 3b, pp. 227-296). London: Academic Press.

Venken, K. J., & Bellen, H. J. (2005). Emerging technologies for gene manipulation in Drosophila melanogaster. *Nat Rev Genet, 6*(3), 167-178. doi:10.1038/nrg1553

Venken, K. J., & Bellen, H. J. (2007). Transgenesis upgrades for Drosophila melanogaster. *Development, 134*(20), 3571-3584. doi:10.1242/dev.005686

Venken, K. J., He, Y., Hoskins, R. A., & Bellen, H. J. (2006). P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science, 314*(5806), 1747-1751. doi:10.1126/science.1134426

Wilson, C., Bellen, H. J., & Gehring, W. J. (1990). Position effects on eukaryotic gene expression. *Annu Rev Cell Biol, 6*, 679-714. doi:10.1146/annurev.cb.06.110190.003335

Wimmer, E. A. (2005). Insect transgenesis by site-specific recombination. *Nat Methods, 2*(8), 580-582. doi:10.1038/nmeth0805-580

Wittkopp, P. J. (2006). Evolution of cis-regulatory sequence and function in Diptera. *Heredity (Edinb), 97*(3), 139-147. doi:10.1038/sj.hdy.6800869

Wittkopp, P. J. (2011). Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of cis- and trans-regulatory differences. *Methods Mol Biol, 772*, 297-317. doi:10.1007/978-1-61779-228-1_18

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., . . . Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science, 326*(5952), 540-544. doi:10.1126/science.1176980

Chapter 3

Evaluating the contribution of *tan* to pigmentation divergence between *D. americana* and *D. novamexicana* using transgenic alleles[2]

Abstract

The pigmentation divergence between the sister species *Drosophila americana* and *D. novamexicana* has been contributed to divergence at the genes *tan* and *ebony.* Here, I investigate the contribution of *tan* on this phenotypic divergence. Using *D. melanogaster* transgenics expressing *D. americana tan*, *D. novamexicana tan*, or a chimeric allele of *tan*, I quantified the pigmentation phenotypes driven by these different *tan* alleles. *D. americana* and *D. novamexicana tan* alleles reliably resulted in significantly different abdominal pigmentation phenotypes in *D. melanogaster*, with the *D. americana* allele producing a significantly darker phenotype. Chimeric alleles of *tan* tested the effect of noncoding sequence, specifically intron 1, on pigmentation. This region of *tan* was previously mapped and shown to contribute to pigmentation divergence between species. Replacing the 5'half of intron 1 in a *D. americana tan* allele with the 5' half intron 1 sequence from *D. novamexicana* resulted in a significant lightening in abdominal pigmentation. These results suggest that the 5' half of *tan* intron 1 contains at least one nucleotide that contributes to pigmentation differences between *D. americana* and *D. novamexicana.* Future analyses can test individual divergent nucleotides for

[2] Research presented in this chapter was done with contributions from Abigail Lamb.

effect on pigmentation to identify the specific genetic change(s) contributing to pigmentation divergence between *D. americana* and *D. novamexicana*.

Introduction

Understanding the often complex relationship between genotype and phenotype is a fundamental goal in biology, and one that remains unresolved. Pigmentation serves as a model trait for investigating this relationship, particularly in *Drosophila*, which has a relatively well-characterized pigmentation pathway (Kopp, 2009; Massey & Wittkopp, 2016; Takahashi, 2013; True, 2003; Wittkopp, Carroll, & Kopp, 2003). Body color in *Drosophila* is a polygenic phenotype with documented cases of genes involved in pigment synthesis (*yellow, ebony, tan,* etc.) and developmental regulators (*bric-a-brac* (*bab*), *abdominal-B* (*Abd-*B), *doublesex* (*dsx*), *Distal-less* (*Dll*), and *Engrailed* (*en*)) of these genes contributing to pigmentation divergence within and between species (Arnoult et al., 2013; Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Jeong, Rokas, & Carroll, 2006; Kopp, Duncan, Godt, & Carroll, 2000; Williams et al., 2008). When specific genetic changes contributing to pigmentation divergence have been identified in *Drosophila*, the changes have always been in noncoding sequences and impacted *cis*-regulatory function (Massey & Wittkopp, 2016). However, expanding the number of cases in which both the specific genetic changes and mechanism through which they influence phenotypic diversity will shed light on if these trends are a reality of the underlying biology of phenotypic evolution or a consequence of the limited number of cases in which both the genetic change and mechanism have been identified.

Drosophila americana and *Drosophila novamexicana* are sister species that diverged from one another approximately ~400 thousand years ago (Morales-Hojas, Vieira, & Vieira, 2008). Since that time, a dramatic difference in pigmentation has developed between the two

species (Figure 3.1); *D. novamexicana* has a derived, lightly pigmented phenotype while *D. americana* and the other members of the *virilis* subgroup have ancestral, dark pigmentation (Wittkopp et al., 2009).

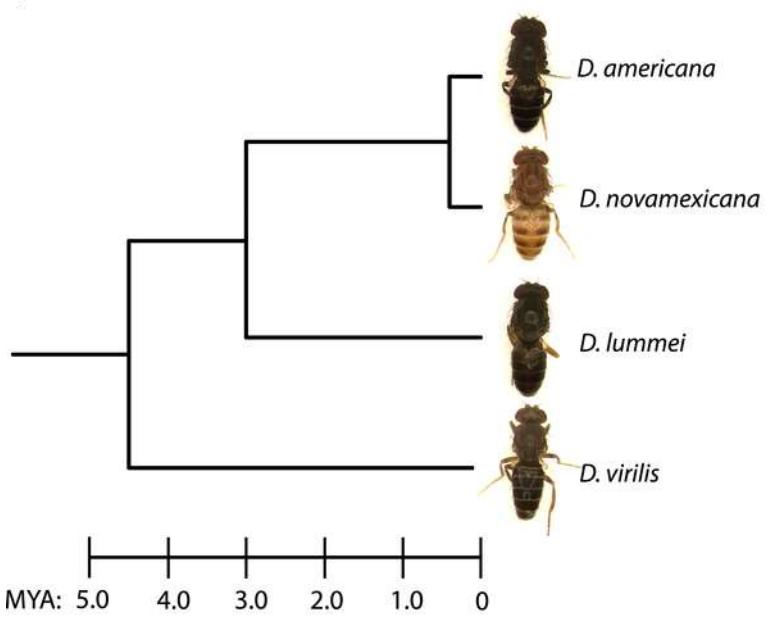

Figure 3.1 Pigmentation in the virilis group of *Drosophila*. Body color for members of the virilis group are shown. *D. americana* has a recently evolved, derived, and dramatically distinct pigmentation phenotype compared to other members of the virilis group. Figure adapted from Cooley et al. 2012.

Previous efforts identified sequences linked to *tan* and *ebony* as responsible for the majority of pigmentation divergence between *D. americana* and *D. novamexicana* (Wittkopp et al., 2009). *tan* and *ebony* catalyze opposite reactions in the melanin pigmentation pathway in *Drosophila* leading to brown and yellow pigment formation, respectively (Figure 3.2). *tan* and *ebony* are differentially expressed during late pupal development consistent with species pigmentation (Cooley, Shefner, McLaughlin, Stewart, & Wittkopp, 2012). Specifically, the darkly pigmented *D. americana* expresses *tan*, which contributes to brown pigment formation, more highly than *D. novamexicana.* The converse was true for *ebony*, which contributes to

yellow pigment formation and showed higher expression in the lightly pigmented *D. novamexicana.* Both *cis*- and *trans*-regulatory divergence contribute to differential expression of *tan* and *ebony* between species (Cooley et al., 2012).

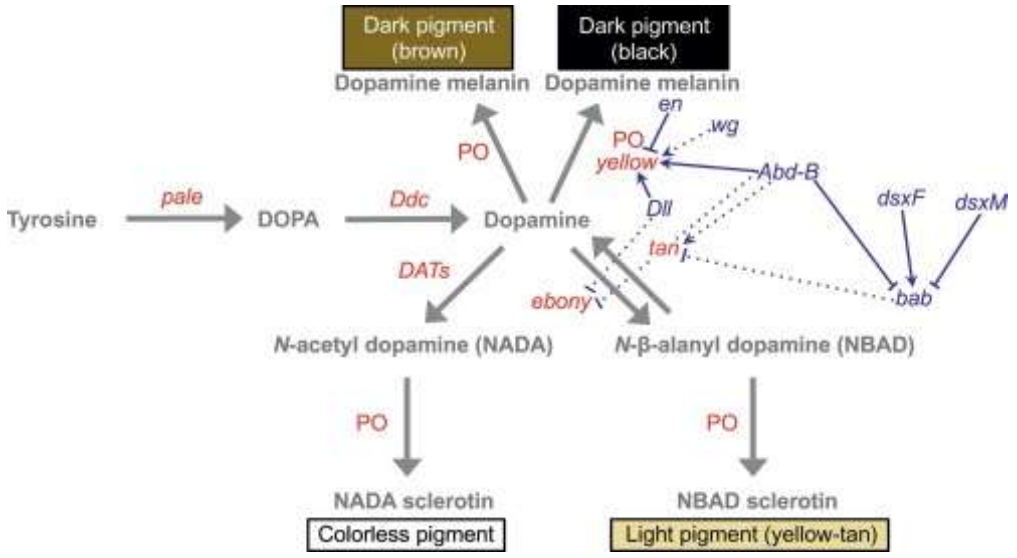

Figure 3.2: Pigmentation biosynthesis in *Drosophila.* From Massey and Wittkopp, 2016. Genes that are part of the pigmentation biosynthesis pathway are shown in red. Genes that are involved in the regulation of pigment development are shown in blue; both direct (solid) and indirect (dashed) regulators are connected with blue lines to the gene(s) they regulate; pointed arrow connections indicate a positive regulatory interaction (i.e. activation) and blunt connection indicate a negative regulatory interaction (i.e. repression). Metabolites are shown in grey. Grey arrows indicate direction of enzymatic reactions in the pathways.

tan is in a freely recombining region of the genome which allowed fine-scale genetic mapping to identify a 2.7-kb region of *tan* that contributes to, but does not entirely explain, *tan*'s effects on pigmentation divergence between species (Wittkopp et al., 2009). The region contains only noncoding fixed differences between species, leading to the hypothesis that noncoding changes in this region, specifically within intron 1, contribute to the pigmentation divergence between *D. americana* and *D. novamexicana*, likely by altering expression of *tan* through changes in *cis*-regulatory elements.

To better understand how changes in *cis*-regulatory sequence can impact phenotypes, I aimed to further refine the 2.7-kb region to identify specific nucleotide changes contributing to pigmentation divergence between *D. americana* and *D. novamexicana*. To this end, transgenic *D. melanogaster* lines were produced by integrating *D. americana tan*, *D. novamexicana tan*, or a chimeric *tan* allele. Each *tan* allele was inserted at the same position in the *D. melanogaster* genome to minimize genomic position effects. Furthermore, the genomic position selected has been shown to capture differences in abdominal pigmentation driven by *D. americana tan* and *D. novamexicana tan* in *D. melanogaster* (John, Sramkoski, Walker, Cooley, & Wittkopp, 2016). Using this transgenic approach, I saw pigmentation differences between transgenic alleles consistent with pigmentation divergence between *D. novamexicana* and *D. americana*, as well as gain some insight into the contributions of *tan* intron 1 to this pigmentation divergence. Here I show: 1) transgenic *tan* alleles, including chimeric *tan* constructs, drive differential pigmentation in a *D. melanogaster* host, 2) noncoding changes in *tan* are sufficient to alter pigmentation in this transgenic analysis suggesting a role for these noncoding sequences in the functional differences between *D. americana* and *D. novamexicana tan* alleles, and 3) the effect of the noncoding changes on pigmentation may be altered by the functional status of *yellow* (a gene involved in dark pigment production).

Materials and Methods

Generation of Transgenic Flies

D. americana, *D. novamexicana* and chimeric *tan* transgenes were constructed using Recombineering (https://redrecombineering.ncifcrf.gov/). Specifically, a targeting plasmid for each species was created by PCR amplifying ~500bp “homology arms” from the genes flanking

tan (HMR and CG7039). The homology arms were connected with an XhoI restriction site between them using PCR sewing. The resulting DNA fragment was inserted into the AscI site of a piggyBAC plasmid (Horn & Wimmer, 2000) using AscI restriction sites that were added to homology arms during initial PCR amplification. The piggyBAC plasmid utilized contains Pax6-EGFP, an eye-expressing green fluorescence marker used to detect successful integration (Horn & Wimmer, 2000) and was modified to include an attB site used for ΦC31-mediated transformation (Groth, Fish, Nusse, & Calos, 2004) at the XbaI site in the plasmid backbone. The 7kb targeting vectors for *D. americana* and *D. novamexicana tan* were independently constructed and sequenced confirmed.

The targeting vectors were linearized with XhoI, gel purified, dephosphorylated, and electroporated into SW102 cells (Warming, Costantino, Court, Jenkins, & Copeland, 2005) with the appropriate BAC containing an edited *tan* allele from each species or a chimeric *tan* allele. SW102 cells contain all necessary components for recombineering (Warming et al., 2005). Relative to the initial BACs used (DA_ABa0020L7 for *D. americana* and DN_Ba0024C15 for *D. novamexicana*), a single amino was changed in each species' allele to replace a rare polymorphism with the most common allele found in the species. Specifically, a A at position 174 in exon 5 was changed to a T nucleotide resulting in a Q190H change in *D. americana* and a C at position 47 in exon 7 was changed to a A nucleotide resulting in a P269T change in *D. novamexicana*. These changes were made using a two-step recombineering protocol (Warming et al., 2005) where the selectable marker *galK* is recombined into a specific location and subsequently replaced by the desired sequence. These rare polymorphisms were found to have no visible effect on pigmentation intensity when the Tan proteins were overexpressed using the GAL4-UAS system (unpublished). Chimeric alleles were constructed in a similar fashion and

with the rare polymorphisms replaced. The chimeric *tan* alleles contain portions of noncoding sequence within intron 1 that have been swapped between species. Specifically, the following chimeric alleles were constructed: 1) *D. americana tan* with the 5' half of intron 1 replaced with *D. novamexicana* sequence (A[N_5'_intron1]), 2) *D. americana tan* with the intron 1 replaced with *D. novamexicana* sequence (A[N_intron1]), and 3) *D. novamexicana tan* with the 5' half of intron 1 replaced with *D. americana* sequence (N[A_5'_intron1]) (Figure 3.3). The intron 1 breakpoint for the 5' half occurs in a conserved region of sequence approximately 1.2kb into intron 1 such that the 3' most SNP included in the 5' half is #2296 as designated in Wittkopp et al. 2009.

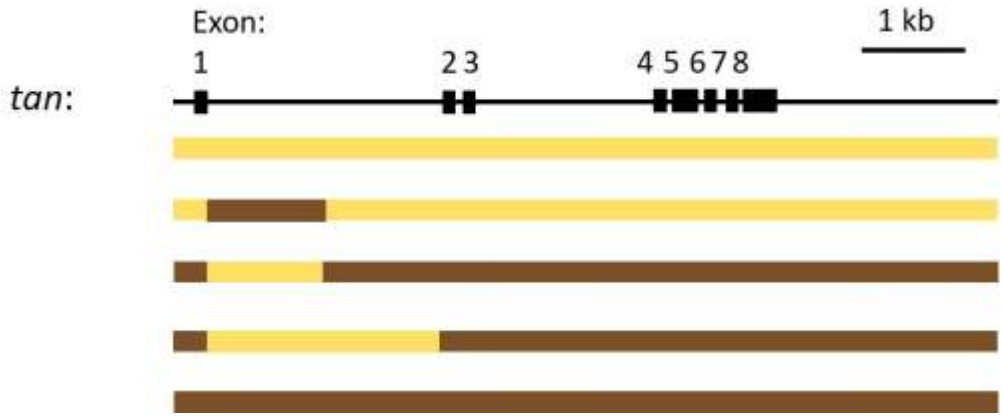

Figure 3.3: Schematic representations of *tan* transgenic alleles. *tan* sequence drawn approximately to scare with black boxes representing exons and black lines representing intronic and intergenic noncoding sequences. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. Constructs from top to bottom are: *D. novamexicana*, *D. novamexicana* with *D. americana* 5' half intron 1 (N[A_5'_half_intron1]), *D. americana* with *D. novamexicana* 5' half intron 1 (A[N_5'_half_intron1]), *D. americana* with *D. novamexicana* intron 1 (A[N_intron1]), and *D. americana*.

After recombineering, individual colonies that grew on selective media (Amp+) were screened via PCR and diagnostic restriction digests for the presence of *tan*. A single positive clone from each *tan* allele was sequence confirmed. The resulting 14kb *tan* transgenes contain 4.1kb of sequence 5' of *tan* and 3.6kb of sequence 3' of *tan* in addition to all exonic and intronic sequences of *tan*.

GenetiVision (Houston, Texas) performed all DNA preparations and embryo injections into a *D. melanogaster* host according to their standard protocols (http://www.genetivision.com/). The *D. melanogaster* host genotype carried a transgene on the X-chromosome using the *vasa cis*-regulatory sequences to express the ΦC31 integrase specifically in the germ-line, a single attP site located on the third (86Fb – BDSC #24749) chromosome, and an eye-expressing red fluorescent protein (RFP) as a visible marker for the landing site (Bischof, Maeda, Hediger, Karch, & Basler, 2007). Transformant flies (expressing green fluorescent protein in their eyes) were used to establish lines homozygous for each transgene in a *D. melanogaster* background carrying loss-of-function mutations in the X-linked genes *tan, yellow,* and *white*. The mutant *D. melanogaster tan* allele allowed us to test for partial rescue of the *tan* mutant pigmentation phenotype by the heterologous *tan* alleles contained in the transgenes; the *yellow* mutant allele reduced the amount of black pigment present in these flies, providing a more sensitive assay for changes in abdominal pigmentation caused by the transgenes; and the *white* allele allowed for easier visualization of the eye-expressing fluorescent transformation marker.

Generation of yellow *mutant in* tan, white *mutant genetic background*

To test the effect of *yellow* on the pigmentation driven by the chimeric *tan* alleles, we sought two lines of *D. melanogaster* with genetic backgrounds containing minimal genetic differences besides the functional status of *yellow*. We also required a mutant copy of *tan* and *white* for the reasons mentioned previously. Since two such lines did not exist, we utilized the CRISPR/Cas-9 system to create a *yellow* mutant in a *tan, white* mutant background obtained from J. True (T20A deletion line).

A 20-bp region in *yellow* exon 1 (GCGATATAGTTGGAGCCAGC) was targeted. This sequence was evaluated for potential off-target effects using the CRISPR Optimal Target Finder (http://tools.flycrispr.molbio.wisc.edu/targetFinder/index.php) and was found to be a strong candidate for specific targeting. The target sequence was cloned into the pCFD3 guide RNA expression plasmid (Port, Chen, Lee, & Bullock, 2014) and injected into approximately 300 embryos from the *tan, white* mutant *D. melanogaster* line. The injection mix contained: 500ng/uL guide RNA plasmid and 500ng/uL pBS-Hsp70-Cas9 plasmid (Gratz et al., 2014). Of the injected embryos, 52 adult survivors emerged and were used to set 1:1 sibling crosses. Three of the 27 crosses set produced some males with no black pigmentation, indicating nonfunctional *yellow*. These mutant *yellow* males were crossed to their female siblings to produce homozygous *yellow* mutant female offspring (*yellow* is located on the X chromosome). These offspring were used to establish homozygous *yellow* mutant lines. Molecular screening to identify the specific lesion in *yellow* has not been performed, however, it is suspected that it is either an indel or frameshift mutation since these are common in non-homologous end joining CRISPR mutants.

Drosophila husbandry, collection, and abdominal cuticle dissection

For each line to be analyzed, virgin females were mated with males from the same homozygous *tan* transgenic line on standard yeast-glucose media at temperatures ranging from 20-25°. The exception to this was for the functional vs nonfunctional *yellow* analysis where virgin females from either a *yellow, tan* mutant line or a *tan* mutant line were mated to males from each homozygous *tan* transgenic line to yield offspring hemizygous for the *tan* transgene. For all crosses, parents were removed from the vials upon formation of pupae and the offspring were

allowed to continue development. Male offspring were collected 0-1 days post-eclosion and aged to 7-8 days. Flies were stored in 10% glycerol in ethanol prior to dissection.

To harvest abdominal cuticles, 7-8 day old males were removed individually from the 10% glycerol in ethanol solution and placed on a glass slide. Using a razor blade, the abdomen was separated from the rest of the body then cut along the lateral edge parallel with the anterior-posterior axis. The dorsal half of the abdomen was soaked overnight in a solution of phosphate buffered saline (PBS; 1.4 M NaCl, 27 mM KCl, 190 mM Na_2HPO_4, 18 mM KH_2PO_4, adjusted to pH 7.4 with 1 M HCl). After soaking overnight, a single dorsal half of abdominal cuticle was removed from the PBS and placed on a glass slide, dorsal (cuticle) side down. Using forceps, the abdominal cuticle was cleared of any remaining debris. The cleaned cuticle was then mounted dorsal side up in polyvinyl alcohol (PVA) mounting media (BioQuip) on a clean glass slide, covered with a coverslip, and the coverslip was sealed with clear nail polish. This process was repeated for all genotypes analyzed in a given experiment. To minimize effects of any day-to-day differences in dissections, all genotypes were dissected during each dissection session.

Image collection and processing

Dissected abdominal cuticles were imaged in a single session using a Leica MZ6 microscope and Scion (CFW-1308C) camera operated via TWAIN driver in Adobe Photoshop. Magnification was set between 3-4 with ring light illumination at 70-80%. At the beginning of each experimental imaging session, auto white balance (AWB) was used and the resulting settings were not changed throughout the imaging session. Imaging was conducted at night to minimize changes in ambient lighting during image collection. Images were taken slide-by-slide (2 cuticles/slides, cuticles imaged individually) with samples arranged such that no more than two

cuticles from the same line were imaged consecutively. A "reference" image of the same dissected cuticle was taken at standard intervals in each experiment to allow us to evaluate the consistency of the image collection, processing, and analysis pipeline throughout the imaging session. Each experimental data set was imaged independently resulting in slight different parameters; however, the procedure outlined here were utilized in all experiments.

For each experiment, all images were compiled into a single document in Adobe Photoshop and the "Levels" function was used to adjust the color of all images simultaneously so that the images more closely matched the cuticle appearance visible by eye. Adjusting all images simultaneously ensures that an identical color adjustment was applied to all photos.

Quantifying pigmentation

Using ImageJ (Rasband, 1997-2016), the area of dorsal abdominal cuticle known as abdominal tergite 4, or "tergite A4", (Figure 3.4) was manually selected using the polygon tool, excluding any regions containing cracks, holes, or overlapping regions. Measurements of pixel intensity (area, mean, standard deviation, mode, min, max, and median) were taken for each selection. In some experiments, the area of dorsal abdominal cuticle known as abdominal tergite 5 (tergite A5) was also measured. These results were compiled into a single Microsoft Excel spreadsheet where allele (no transgene control, *D. americana, D. novamexicana, D. amer.* + 5' half *D. nova.* intron 1, *D. amer.* + *D. nova.* intron 1, *D. nova*. + 5' half *D. amer.* intron 1) identity was added to each entry. Since ImageJ quantifies pigmentation (pixel intensity of a grayscale image) on a 0-255 scale (dark-light), I subtracted the reported pixel intensity from 255 so that darker cuticles had a higher pigmentation score. Each experimental data file was then saved as a .csv file for statistical analysis in R.

Figure 3.4: Pigmentation was measured in the fourth abdominal tergite (A4), indicated with black brackets. In some experiments, pigmentation was also measured in tergite A5, the abdominal segment below A4.

Data analysis

Median pigmentation intensity of tergite A4 and/or tergite A5 for each sample reported by ImageJ was analyzed using R v3.2.5 (RCoreTeam, 2016). Median pigmentation was chosen for analysis instead of mean pigmentation intensity to minimize the impact of outlier (excessively white or black) pixels. Pairwise t-tests using unpooled standard deviations were performed on the pigmentation medians to identify which comparisons among *tan* alleles were statistically significant. Statistical significance was assessed using p-values adjusted for multiple testing by the Benjamini and Hochberg method (Benjamini & Hochberg, 1995) as implemented in the pairwise.t.test function in R.

Results

Transgenic tan *alleles are used to evaluate the impact of noncoding sequences on pigmentation*

After identifying *tan* as contributing to divergent pigmentation between *D. americana* and *D. novamexicana*, further recombination mapping identified a 2.7 kb region of *tan* that contributes to pigmentation divergence between species (Wittkopp et al., 2009). Intron 1 of *tan* contains the only fixed sequences differences between species in this region. I aimed to further

refine this region to identify specific sequences that impact pigmentation. To do so, I tested for pigmentation differences driven by *tan* transgenes. *D. americana* and *D. novamexicana tan* transgenes were constructed previously (Wittkopp et al., 2009), as well as chimeric *tan* alleles that substitute portions of the 2.7kb region between species alleles. All *tan* transgenes were transformed into the *D. melanogaster* genome and quantified their effect on pigmentation. Constructs were examined at a genomic location on the third chromosome (86Fb) where pigmentation differences between driven by *D. americana* and *D. novamexicana tan* alleles are detected (John et al., 2016).

In this section, a collection of experiments will be presented in chronological order. Since each experiment informed and motivated subsequent experiments, a chronological approach to presenting these data is fitting. Additionally, each experiment is independent, such that data can only be compared within a single experiment and cannot be combined between experiments. This is due to variable conditions (e.g. rearing temperature, nutrient composition of the food, and imaging lighting conditions) between experiments that can have a systematic impact on the pigmentation phenotypes seen between different experiments, but are assumed to impact all genotypes in a single experiment equally. Throughout the chapter the *tan* constructs used will be abbreviated as follows: A = *D. americana*, N = *D. novamexicana*, A[N_5'_intron1] = *D. americana* with 5' half *D. novamexicana* intron 1, A[N_intron1] = *D. americana* with *D. novamexicana* intron 1, and N[A_5'_intron1] = *D. novamexicana* with 5' half *D. americana* intron 1. These experiments test the hypothesis that sequence in intron 1 of *D. novamexicana tan* contribute to the derived pigmentation phenotype of that species; this hypothesis is informed from previous mapping studies (Wittkopp et al., 2009) and leads to the expectation that flies

carrying *tan* alleles with *D. novamexicana* sequence within *tan* intron 1 will be lighter in pigmentation than those with *D. americana* sequence in the same region.

In the earliest experiment, only N[A_5'_intron1] (24, n=12) and two independent lines of A[N_5'_intron1] (209, n=8; 211, n=9) were available for analysis. When compared to the reference *yellow, tan* mutant (n=7), all three lines significantly increased pigmentation in tergite A5 compared to the reference line, indicating that the transgenes were being expressed and the Tan protein was functional (Figure 3.5). Although the transgene carrying the 5' half of *tan* intron 1 from *D. americana* (N[A_5'_intron1]; line 24) had a higher median pigmentation score (indicating darker pigmentation), there was no statistically significant difference between the two different transgenic constructs. The lack of a significant difference between these specific constructs is surprising given that they differ throughout most the *tan* locus, including within intron 1 which was previously implicated in pigmentation divergence (Wittkopp et al., 2009). Possible explanations for the lack of pigmentation difference detected include the relatively small sample size and a lack of pigmentation measurements from tergite A4, which showed more substantial differences in pigmentation in later experiments. The addition of full *D. americana* and/or *D. novamexicana tan* constructs in later experiments also helps identify if the phenotypes driven by the constructs tested represent intermediate phenotypes or if this set of *tan* transgenics in *D. melanogaster* does not capture the pigmentation differences driven by *tan* between *D. americana* and *D. novamexicana*.

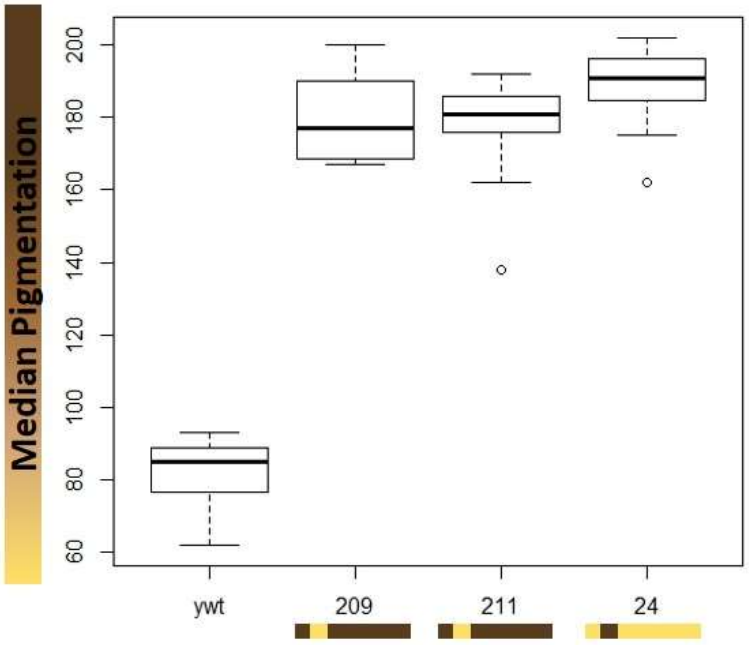

Figure 3.5: Box plots summarizing the pigmentation phenotypes observed for each genotype. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. ywt is the *yellow, white, tan* mutant control. 209, 211 are A[N_5'_half_intron 1]. 24 is N[A_5'_half_intron1].

In the next experiment, the following *tan* transgenes were examined in *D. melanogaster*: *D. americana* (97; n=7), two independent lines of A[N_intron1] (39, n=2; 28; n=6), and two independent lines of N[A_5'_intron1] (55, n=8; 59; n=6, respectively) were analyzed. In both tergite A4 (data not shown) and A5 (shown), I again saw that the N[A_5'_intron1] flies had darker median pigmentation, but these differences were not statistically significant; nor were any other differences between alleles (Figure 3.6). As in the previous experiment, a possible explanation for the inability to detect significant pigmentation differences is low power because of a small sample size. Additionally, in both this experiment and the previous experiment the *D. novamexicana tan* allele was not available for analysis; this construct hypothetically provides the lightest pigmentation phenotype and therefore would help contextualize the results of the other constructs tested. Overall, however, these results suggest that the *D. novamexicana* intron 1 may lighten pigmentation as seen by the relatively lighter median pigmentation seen in both lines of

A[N_intron1] tested. Surprisingly, N[A_5'_intron1] shows a possible increase in median pigmentation compared to *D. americana*. This is surprising due to evidence that other sequences in *tan* contribute to the pigmentation differences between species (Wittkopp 2009); these sequences would be present in the *D. americana* construct, but not in N[A_5'_intron1]. However, if these results are valid, they suggest that at least one causative nucleotide(s) in intron 1 reside in the 5' half of *tan* intron 1.

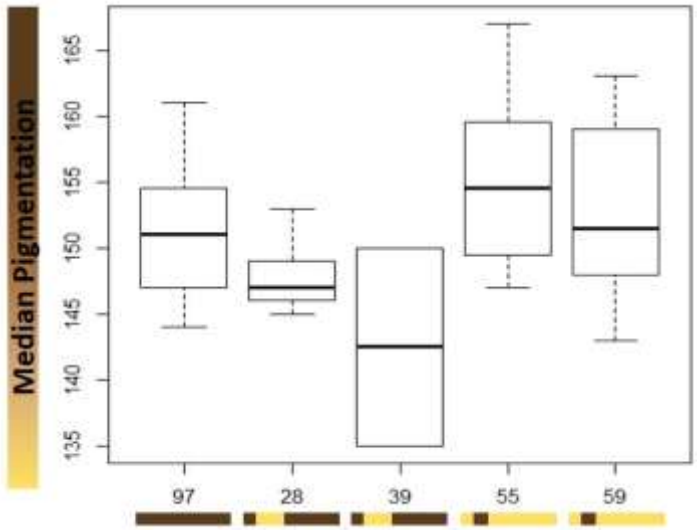

Figure 3.6: Box plots summarizing the pigmentation phenotypes observed for each genotype. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 97 is *D. americana*. 28, 39 are A[N_intron 1]. 55, 59 are N[A_5'_half_intron1].

Due to lack of statistically significant differences in pigmentation detected in the previous two experiments, the next experiment was conducted using only two independent lines of *D. americana* (97, n=10; 98; n=10) and two independent lines of *D. novamexicana* (117, n=111; 118; n=12). If these two alleles do not show significant differences in pigmentation at the genomic location tested, evaluating additional chimeric *tan* alleles is unproductive. Pigmentation was quantified in both tergite A4 and A5. Data obtained from tergite A4 showed that both alleles significantly increased pigmentation relative to a *yellow*, *tan* mutant (Figure 3.7A). Additionally,

significant differences in pigmentation were detected between *D. americana* and *D. novamexicana tan* alleles in a matter consistent with pigmentation differences seen between species. Surprisingly, the two lines of *D. novamexicana tan* were also significantly different from one another, although both were still significantly lighter than the *D. americana tan* lines. Analyzing pigmentation in tergite A5 yielded different results; specifically, pigmentation was not significantly different between *tan* constructs, despite all constructs increasing pigmentation significantly from the *yellow, tan* background (Figure 3.7B).

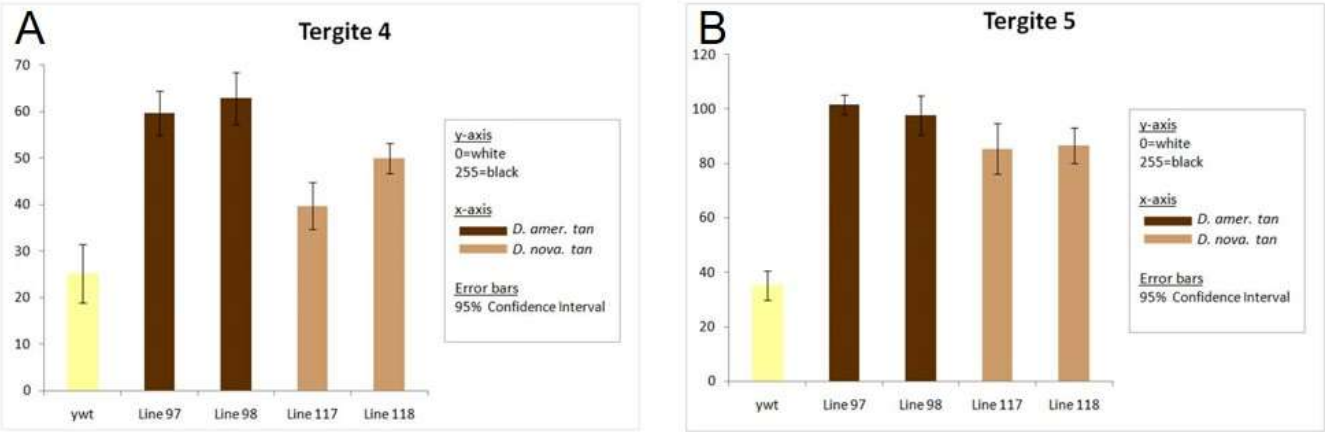

Figure 3.7. Median pigmentation (y-axis) for each genotype tested. Error bars represent 95% confidence intervals. ywt is the *yellow, white, tan* mutant control. 97, 98 are *D. americana*. 117, 118 are *D. novamexicana*. Results for measurements in tergite A4 (A) and tergite A5 (B) are shown.

Next, I evaluated *D. americana* (97, n=12), *D. novamexicana* (118, n=12), and N[A_5'_intron1] (59, n=12). When measured from tergite A4, all constructs are significantly different from one another with N[A_5'_intron1] (59) showing the lightest pigmentation and *D. americana* showing the darkest pigmentation (Figure 3.8). This result is surprising since N[A_5'_intron1] had shown the darkest pigmentation phenotype in previous experiments. When measured in tergite A5, N[A_5'_intron1] (59) remained significantly lighter than the other lines, however, I no longer detected a significant difference in pigmentation between *D. americana* and *D. novamexicana* (Figure 3.9).

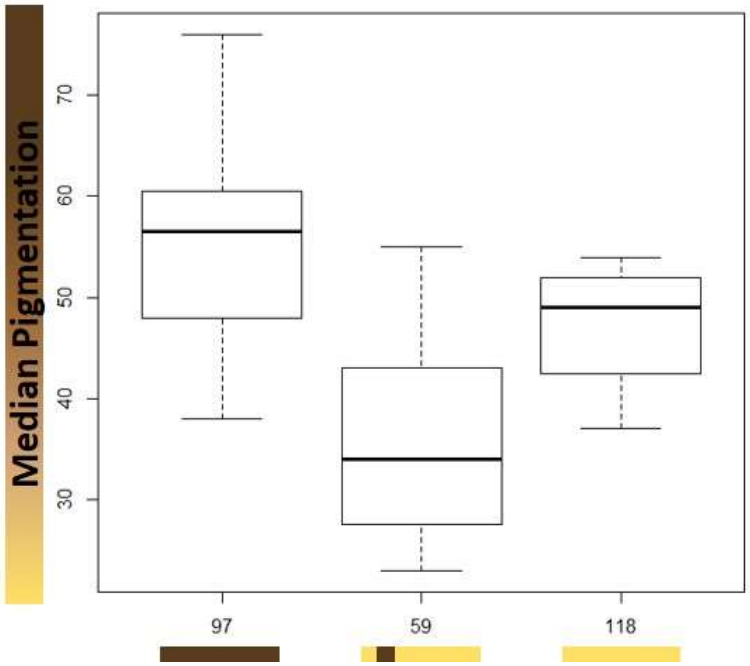

Figure 3.8: Box plots summarizing the pigmentation phenotypes observed in tergite A4. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 97 is *D. americana*. 59 is N[A_5'_half_intron1]. 118 is *D. novamexicana*.

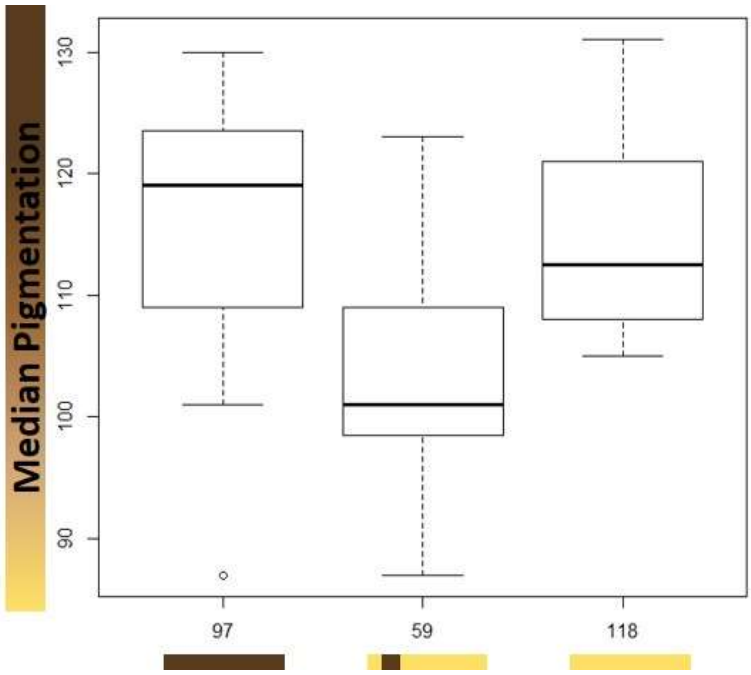

Figure 3.9: Box plots summarizing the pigmentation phenotypes observed in tergite A5. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 97 is *D. americana*. 59 is N[A_5'_half_intron1]. 118 is *D. novamexicana*.

In the next experiment, all available transgenic lines were analyzed, including a control *yellow, tan* mutant line (ywt, n=25), *D. americana* (98, n=32), A[N_5'_intron1] (209, n=28), A[N_intron1] (28, n=28), N[A_5'_intron1] (59, n=22), and *D. novamexicana* (120, n=32). When looking at pigmentation in tergite A4, all constructs are significantly different from the *yellow, tan* mutant (Figure 3.10). *D. americana* and A[N_intron1] are not significantly different from one another, but are significantly different from the remainder of the constructs tested. Similarly, *D. novamexicana* and A[N_5'_intron1] are not significantly different from one another, but are significantly different from the remainder of the constructs tested. N[A_5'_intron1] is significantly lighter than all of the other constructs evaluated (expect the *yellow, tan* mutant). This last result, while surprising, agrees with the immediately previous data set (yet not with the initial experiments presented) and suggests that the 5' half of *D. americana* intron 1 contains an element that lightens pigmentation, at least in the *D. melanogaster* genetic background. Another surprising result is that A[N_intron1] is indistinguishable statistically from *D. americana* while *D. americana* with only the 5' half of *D. novamexicana* intron 1 is significantly lighter and indistinguishable from *D. novamexicana*. Our expectation would be that if the 5' half of *D. novamexicana* intron 1 is sufficient to lighten pigmentation then the full *D. novamexicana* intron 1 should also lighten pigmentation. A possible explanation is that there exists an element within the 3' half of *D. novamexicana* intron 1 that darkens pigmentation. As in previous experiments, analyzing pigmentation in tergite A5 did not produce significant differences among any of the *tan* alleles tested, although all constructs significantly increased pigmentation from the *yellow, tan* mutant background (Figure 3.11). Tergite five appears to be qualitatively darker in pigmentation compared to tergite four, regardless of *tan* identity, which may reduce our ability to detect differences in pigmentation driven by the *tan* transgenes.

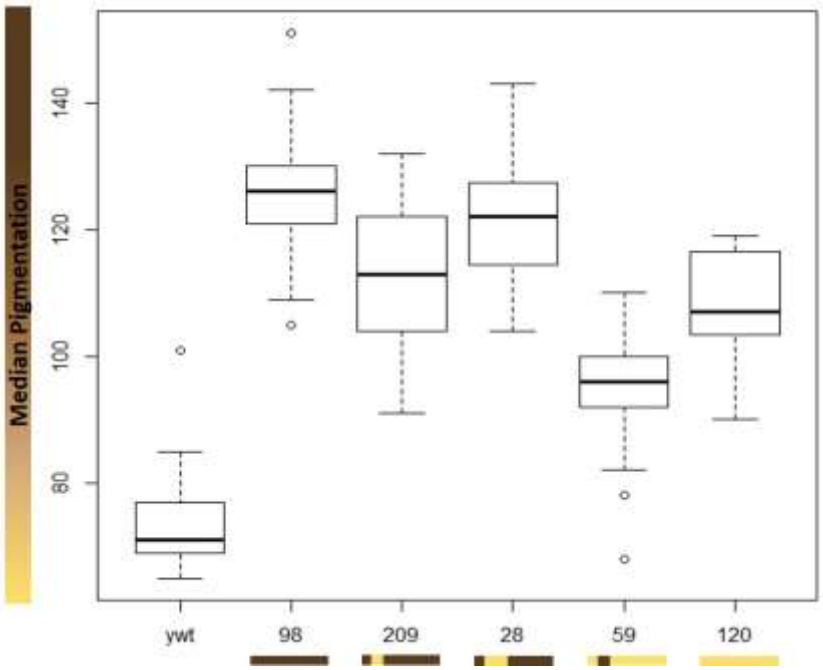

Figure 3.10: Box plots summarizing the pigmentation phenotypes observed in tergite A4. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 98 is *D. americana*. 209 is A[N_5'_half_intron1]. 28 is A[N_intron1]. 59 is N[A_5'_half_intron1]. 120 is *D. novamexicana*.

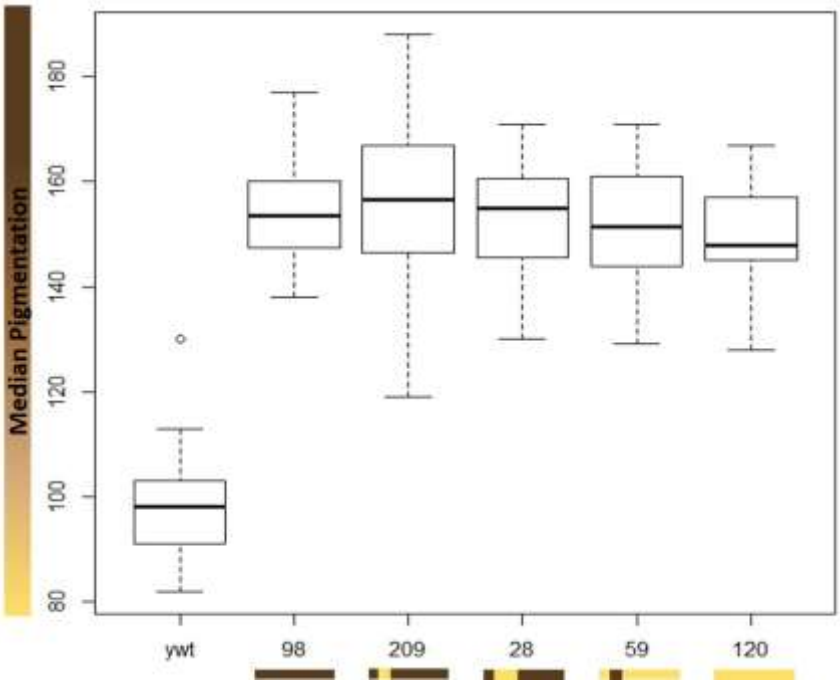

Figure 3.11 Box plots summarizing the pigmentation phenotypes observed in tergite A5. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 98 is *D. americana*. 209 is A[N_5'_half_intron1]. 28 is A[N_intron1]. 59 is N[A_5'_half_intron1]. 120 is *D. novamexicana*.

Using different, independent lines from each of the constructs analyzed in the previous experiment, I repeated the experiment to determine if the results were reproducible. This experiment included *D. americana* (96; n=10), *D. novamexicana* (117, n=10), A[N_5'_intron1] (211, n=10), A[N_intron1] (29, n=10), N[A_5'_intron1] (24, n=10). As in the previous experiment, *D. americana* and A[N_intron1] are not significantly different from one another, but are significantly different from the remainder of the constructs tested (Figure 3.12). However, in this data set, *D. novamexicana* and N[A_5'_intron1] are clustered together, rather than *D. novamexicana* and A[N_5'_intron1] as in the previous experiment. Furthermore, A[N_5'_intron1] is significantly lighter than both *D. americana* and A[N_intron1] and significantly darker pigmentation than *D. novamexicana* and N[A_5'_intron1]. The comparison between A[N_5'_intron1] and A[N_intron1] is again surprising in this experiment since A[N_intron1] contains all of the sequence from A[N_5'_intron1] plus additional *D. novamexicana* sequence, which is hypothesized to lighten pigmentation.

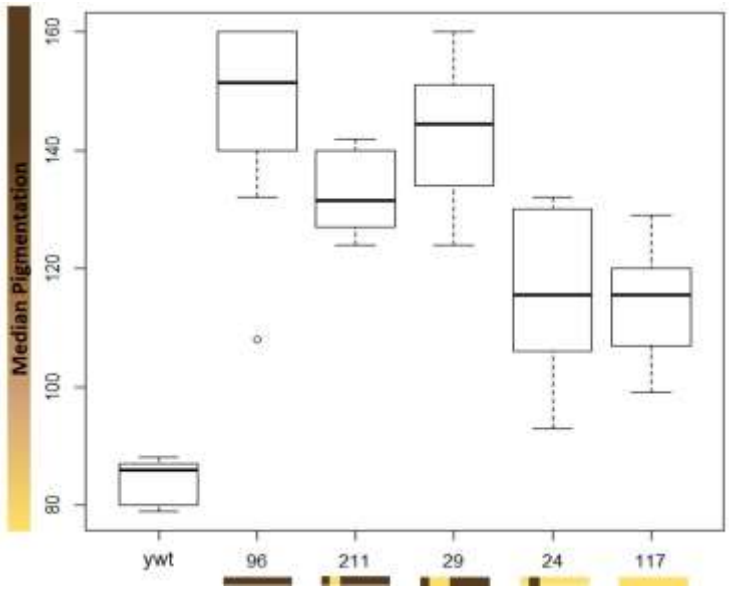

Figure 3.12: Box plots summarizing the pigmentation phenotypes observed for each genotype. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 98 is *D. americana*. 209 is A[N_5'_half_intron1]. 28 is A[N_intron1]. 59 is N[A_5'_half_intron1]. 120 is *D. novamexicana*.

While efforts were made to ensure homozygosity prior to dissection and cuticle imaging, the multitude of surprising results suggested that there may be residual heterozygosity in a subset of the lines tested. To create homozygous transgenic lines, I utilized a third chromosome balancer which was put into a background containing an X chromosome with nonfunctional copies of the pigmentation genes *yellow* and *tan*. Balancer chromosomes have a dominant phenotypic marker allowing for identification of heterozygotes, a recessive lethal mutation ensuring that the only homozygotes obtained have two copies of the desired chromosome, and are riddled with inversions that prevent recombination along the chromosome of interest. The third chromosome balancer used in these experiments is marked with a mutant bristle phenotype (*Hu*), which can be difficult to accurately identify its presence/absence which may have led to residual heterozygosity in these lines. While *Hu* is reported to have high penetrance (Gompel & Chyb, 2013), in our laboratory we have noticed that the phenotype is often hard to identify and/or only present on one side of the fly. The balancer chromosome used in these studies also contains an *ebony* mutation which darkens pigmentation slightly. The presence of this chromosome, especially if in differing amounts between *tan* transgenic lines, could have contributed to the surprising results throughout the previously presented experiments. To more definitively test this possibility, two experiments were conducted. First, presumably homozygous males from *D. americana* (96), A[N_5'_intron1] (211), A[N_intron1] (29), and *D. novamexicana* (117) were mated to virgin females from the control *yellow, tan* mutant line (ywt). If the males are homozygous, all offspring from this cross are expected to be heterozygous for the transgene and thus have GFP expression in the eye driven by the 3xP3-GFP screening marker in the transgenic construct. However, if there is residual heterozygosity in the *tan* transgenic population, one would expect a mix of flies with and without GFP expression. In these crosses,

both the *D. americana* and *D. novamexicana* lines tested showed a mix of flies with and without GFP expression indicating residual heterozygosity in these lines. In the second experiment, genomic DNA was extracted from a pool of 12 F1 hybrid flies; these F1 hybrids were created by mating *D. novamexicana* (117) to either *D. americana* (96), A[N_5'_intron1] (211), or A[N_intron1] (29). Using pyrosequencing, the relative allele frequency of *D. americana* and *D. novamexicana* was measured. If both lines are homozygous, these alleles are expected to be present in approximately equal amounts with only slight deviations from the expected 1:1 ratio caused by PCR bias. If one of the lines used was heterozygous, a lower proportion of the hetetozygotes' allele in the genomic DNA of the F1 hybrids is expected. A drawback to this design is that if both lines used to create the F1 hybrids have a similar rate of residual heterozygosity, a skew in allele abundance would not be detected; however, taken with the crossing design presented above, clear conclusions regarding residual heterozygosity can be made. In this experiment, there was a clear skew away from the *D. novamexicana* allele in both the A[N_5'_intron1] (211) and A[N_intron1] (29) shown by a nearly 2:1 ratio of *D. americana* to *D. novamexicana* and suggesting that Line 117 had residual heterozygosity (Table 3.1). The results from the *D. americana* x *D. novamexicana* (96x117) F1 hybrids is less clear since it varies only slightly from the expected 1:1 ratio. With the results from the GFP crossing experiment, however, it is possible to conclude that both of these lines have residual heterozygosity, likely to a similar degree.

Table 3.1: Allele quantification in F1 hybrid offspring. Expected ratio of alleles is 1:1, resulting in a amer/nova ratio approximately equal to 1.

tan Exon 8 - Da = T; Dn = C			
	T	C	amer/nova
29x117 F1 gDNA	17.61	8.64	2.04
29x117 F1 gDNA	15.07	6.48	2.33
96x117 F1 gDNA	12.51	15.12	0.83
96x117 F1 gDNA	9.97	10.77	0.93
211x117 F1 gDNA	15.17	8.08	1.88
211x117 F1 gDNA	14.62	7.25	2.02

Taken together, these results suggest that a number of the experiments presented previously contained constructs with unaccounted for residual heterozygosity. The lack of two copies of the *tan* transgenes and the presence of the *ebony* mutation in the balancer line likely impacted pigmentation phenotypes throughout these experiments; the degree to which this influenced results, however, is unclear. Some of the inconsistent results and the variance within and between the experimental data sets is likely attributable to residual heterozygosity. Many of the general trends are likely robust to these influences, especially those with reproducible results throughout multiple experiments; these are described in the Discussion section.

Differences in pigmentation driven by chimeric tan alleles may be altered by the functional status of yellow

In an effort to sensitize the phenotypic background to changes in pigmentation caused by *tan*, all analyses discussed previously were carried out in a background with mutant alleles of the pigmentation genes *yellow* and *tan*. Since *yellow* is involved in the production of dark (black) pigments (Wittkopp, True, & Carroll, 2002; Wittkopp, Vaccaro, & Carroll, 2002), it was thought that this would allow for more subtle darkening in pigmentation caused by *tan* to be detected. The mutant *tan* was chosen so that the only functional *tan* present in the flies was from the *tan* transgene introduced into the genome. In the *yellow, tan* mutant background, both species *tan*

alleles significantly (pairwise t-test $p<0.05$) increased pigmentation relative to the *yellow, tan* mutant, with the *D. americana* allele darkening pigmentation significantly more than the *D. novamexicana* allele (Figure 3.10 and 3.12, amongst others). This result follows expectations based on the pigmentation phenotypes of the species. However, substitution of *D. novamexicana* intron 1 into an otherwise *D. americana tan* allele (A[N_intron1]) did not significantly lighten pigmentation relative to the unaltered *D. americana tan*. While this result may be the effect of residual heterozygosity in the transgenic lines, it is also possible that the phenotypic background, specifically nonfunctional *yellow*, caused the alleles to behave differently than they do in their native context. This hypothesis is based on evidence that *yellow* and *tan* may act in concert to produce dark pigment (Jeong et al., 2008) and therefore re-examined abdominal pigmentation in the presence of functional *yellow*.

In the presence of a functional copy of *yellow*, the *D. americana* and *D. novamexicana tan* alleles differentially darken pigmentation as in the *yellow, tan* mutant background. A[N_5'_intron1] produces pigmentation indistinguishable from *D. novamexicana tan*, as in the *yellow, tan* mutant background (Figure 3.13A). However, with a functional copy of *yellow*, substitution of A[N_intron1] also produces pigmentation indistinguishable from both *D. novamexicana tan* and A[N_5'_intron1]. In the *yellow, tan* background, A[N_intron1] failed to significantly lighten pigmentation relative to *D. americana tan* (Figure 3.13B). These results add further support to the conclusion that noncoding changes in intron 1, specifically in the 5' half of intron 1, contribute to pigmentation divergence between species. Furthermore, the results obtained in the presence or absence of *yellow* provides further evidence that *yellow* and *tan* work together to produce dark pigments.

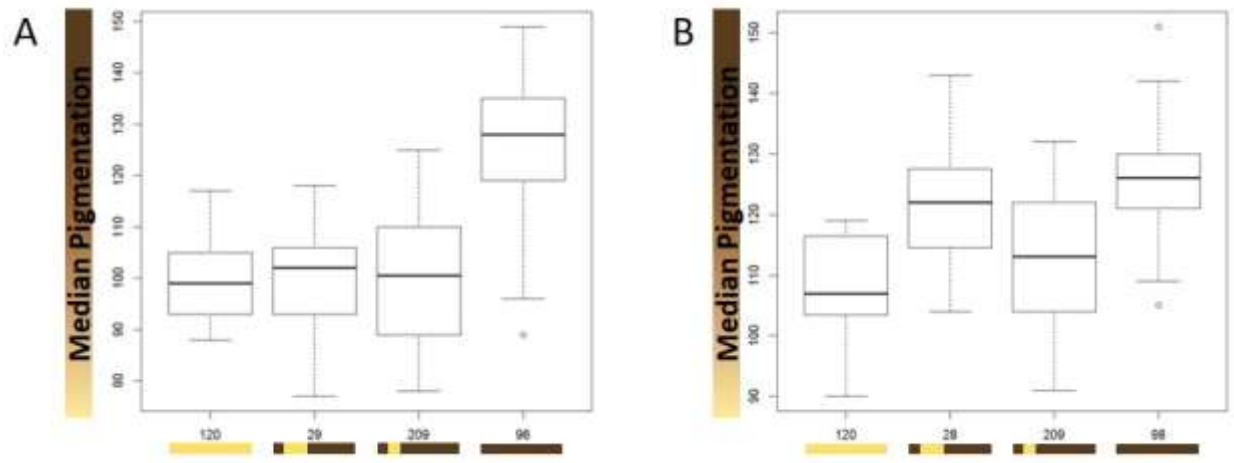

Figure 3.13: Box plots summarizing the pigmentation phenotypes observed for each genotype. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 120 is *D. novamexicana*. 29 is A[N_intron1]. 209 is A[N_5'_half_intron1]. 98 is *D. americana*. (A) Shows results in a *tan, white* mutant background containing a functional copy of *yellow*. (B) Shows results in a *yellow, white, tan* mutant background with a nonfunctional copy of *yellow.*

While the results produced were interesting, several differences existed between this functional *yellow* data set and the previous nonfunctional *yellow* datasets. These include different individuals before the cuticle dissections, imaging, and analysis, as well as differences in the number of transgene copies present; in the nonfunctional *yellow* datasets, the *tan* transgenes were homozygous, however, in the functional *yellow* dataset the *tan* transgenes were hemizygous. The functional *yellow* flies were obtained by crossing virgin females from a *tan* mutant line (with functional *yellow*) to males carrying a *tan* transgene in the *yellow, tan* mutant background. F1 males from this cross inherit a functional *yellow* from the X chromosome of the female, and are heterozygous throughout the rest of the genome, leading to hemizygosity of the *tan* transgene. To rectify these differences, I sought to conduct a more comparable experiment. To do so, each *tan* transgene was crossed into either a functional or nonfunctional *yellow* background such that the transgenes were all hemizygous. Additionally, the functional and nonfunctional *yellow* backgrounds utilized should have minimal differences outside of *yellow* since the nonfunctional

background was made by targeting a portion of *yellow* exon 1 using CRISPR in the functional *yellow* genetic background. All other fly collection, dissection (n=30), and imaging procedures were similar to those used previously.

When pigmentation was analyzed in these hemizygous flies, the overall pigmentation for each line showed greater variability than previous studies. In both the functional and nonfunctional *yellow* background, the full intron 1 from *D. novamexicana* was not sufficient to lighten pigmentation relative to *D. americana*; furthermore, it was not significantly different from the *D. americana* or *D. novamexicana* allele in either case (Figure 3.14). Despite this, the 5' half of *D. novamexicana* intron 1 did significantly lighten pigmentation in the functional *yellow* background; however, in the nonfunctional yellow background it was indistinguishable from *D. americana* and significantly darker than *D. novamexicana*, a result contradictory to all previous experiments. Attempts to refine this data through independent image adjustment for each genetic background (rather than a single adjustment for all images in both genetic background due to the extreme pigmentation differences caused only by the functional status of *yellow*) and different pigmentation quantification methods (i.e. measuring only the area outside of the dorsal midline) failed to yield results from which clearer conclusions could be drawn. Currently, the impact of *yellow* on pigmentation in these *tan* transgenic lines is unclear. Creating homozygous lines in each genetic background may help resolve this uncertainty.

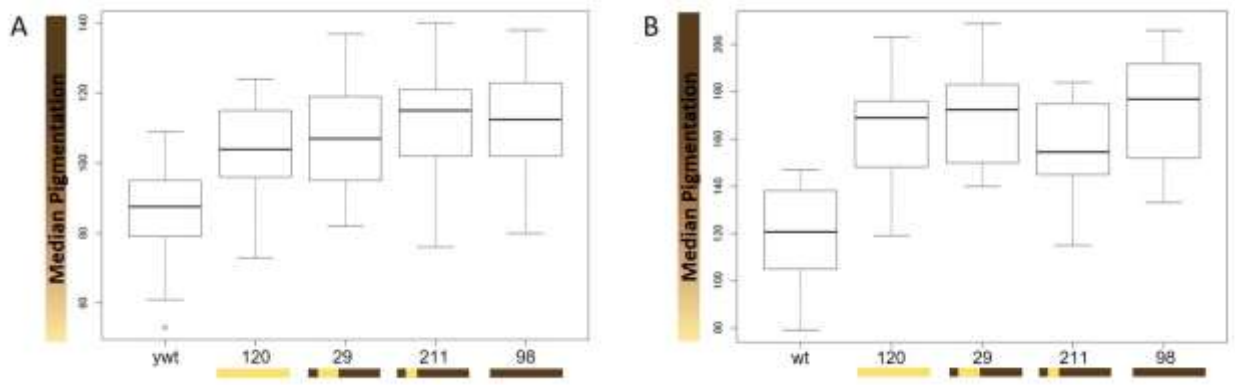

Figure 3.14: Box plots summarizing the pigmentation phenotypes observed for each genotype. The median (center line), first quartile (bottom of box), third quartile (top of box), and ± 1.5 × the interquartile range (whiskers) are shown for each genotype examined. Bars along the x-axis indicate *tan* allelic identity. Brown bars represent *D. americana* sequence. Yellow bars represent *D. novamexicana* sequence. 120 is *D. novamexicana*. 29 is A[N_intron1]. 211 is A[N_5'_half_intron1]. 98 is *D. americana*. (A) Shows results in a *yellow, white, tan* (ywt)mutant background with a nonfunctional copy of *yellow*. (B) Shows results in a *tan, white* (wt) mutant background containing a functional copy of *yellow*.

Discussion

D. americana and D. novamexicana tan alleles reliably produce differential pigmentation in D. melanogaster transgenics (across multiple genomic locations and experimental conditions)

Prior work showed detectable pigmentation differences between *D. americana* and *D. novamexicana tan* alleles across most genomic locations tested (John et al., 2016); however, it did not address the reproducibility of results at a single genomic location. Throughout the experiments presented in the previous section, I measured pigmentation driven by *D. americana* and *D. novamexicana tan* alleles at a single genomic location in five independent experiments over the span of four years. The lines utilized contain the transgenes on the third chromosome (86Fb) in a genomic location that showed significantly different pigmentation driven by *D. americana* and *D. novamexicana tan* alleles previously (John et al., 2016).

In each of the five experiments, pigmentation was measured in tergite A4 of 7-8 day old male flies. Throughout these experiments, the number of flies examined, the temperature of rearing, and the specific transgenic lines used varied slightly. Despite these differences in experimental conditions, in four of five of these experiments, *D. americana* and *D.*

novamexicana tan alleles drove differential pigmentation in tergite A4. In the remaining experiment, the pigmentation difference between *D. americana* and *D. novamexicana tan* alleles narrowly missed a statistical cutoff for significance (p=0.0507). This result suggests pigmentation differences driven by *D. americana* and *D. novamexicana tan* alleles are robust to minor variations in experimental conditions.

Although significant differences detected between species alleles were detected in tergite A4 across multiple experimental conditions, results in tergite A5 (one segment below tergite four) were less consistent. When pigmentation differences are detected within a tergite, the most striking differences occur outside of the dorsal midline stripe in the anterior regions of the cuticle. Tergite A5 appears to be qualitatively darker in pigmentation compared to tergite A4, regardless of *tan* identity, which may reduce our ability to detect differences in pigmentation driven by the *tan* transgenes.

Noncoding changes in tan intron 1 are sufficient to alter pigmentation in D. melanogaster transgenics

While differences in pigmentation between *D. americana* and *D. novamexicana tan* alleles were reliably detected, the impact of intron 1 sequences on pigmentation elude reliable detection or fail to change pigmentation enough to be deemed significant. For instance, the chimeric allele which swaps the full intron 1 sequence from *D. novamexicana* into an otherwise *D. americana tan* background (A[N_intron1]) lightens pigmentation slightly, but not significantly compared to the fully *D. americana tan* allele. Surprisingly, the chimeric allele with only the 5’ half of intron 1 sequence from *D. novamexicana* in an otherwise *D. americana tan* background (A[N_5’_intron1]) does significantly lighten pigmentation compared to the *D.*

americana tan. This suggests that noncoding changes in the 5' half of intron 1 are sufficient to produce changes in pigmentation.

Overall, these transgenic studies suggest that the 2.7kb region of noncoding sequence identified previously contributes to pigmentation divergence between species since chimeric *tan* alleles that change only these sequences can significantly alter pigmentation. Understanding of this system, as well as how noncoding sequences can contribute to phenotypic evolution in general, would benefit from the identification of a specific nucleotide or nucleotides causing differences in pigmentation and the mechanism through which they exhibit their effect. With a more extensive collection of causative nucleotide(s) and mechanism of action in phenotypic evolution, we may eventually be able to better predict what impact, if any, a change in noncoding sequence would have on a particular phenotype.

Despite these conclusions being able to be made using the *D. melanogaster* transgenic system, a more biologically relevant experiment would be to test these chimeric alleles in the native genomic background of *D. americana* and/or *D. novamexicana*. This would allow the noncoding sequences to be evaluated in a regulatory genetic background that they normally exist in. Additionally, this experimental design would preserve any genetic interactions that may exist between other genes in the *D. americana* and/or *D. novamexicana* genome and *tan.* By evaluating the impact of non-coding sequences of *tan* in their native (or at least more closely related) genomic background, more reliable conclusions regarding their effect will be drawn and the effect of individual nucleotides can be evaluated. The advent of technologies like CRISPR broaden the experimental possibilities in non-model species and these avenues of experimentation are being pursued in this system. Alternatively, the results presented may also

suggest that many changes in *tan* and their epistatic interactions with one another contribute to phenotypic divergence between species. In this case, testing multiple noncoding changes at the same time, for instance intron 1 and intron 3 (the other large intron in *tan*), would be more ideal and may show a greater impact on pigmentation than either intron alone.

Acknowledgments

This work would not have been possible without: Lisa Sramkoski and Arielle Cooley for their contributions to the foundation of this work, including generation of *tan* transgenics; Bradley Lankowsky, Wesley McLaughlin, Robert Dikeman, and Elizabeth Walker for assistance in transgenic construction, abdominal cuticle dissections, and general fly husbandry; Abigail Lamb for the initial pigmentation quantification in the functional *yellow* background and generation of the *yellow, tan* mutant used in the experiment testing the effect of the functional status of *yellow*; and members of the Wittkopp lab for helpful discussions and experimental assistance. This work was funded by the National Institute of Health (R01-GM-089736).

References

Arnoult, L., Su, K. F., Manoel, D., Minervino, C., Magrina, J., Gompel, N., & Prud'homme, B. (2013). Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. *Science, 339*(6126), 1423-1426. doi:10.1126/science.1233749

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological, 57*(1), 289-300.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., & Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A, 104*(9), 3312-3317. doi:10.1073/pnas.0611511104

Cooley, A. M., Shefner, L., McLaughlin, W. N., Stewart, E. E., & Wittkopp, P. J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in Drosophila

americana and D. novamexicana. *Evol Dev, 14*(4), 317-325. doi:10.1111/j.1525-142X.2012.00550.x

Gompel, N., & Chyb, S. (2013). *Atlas of Drosophila Morphology: Wild-type and Classical Mutants*.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature, 433*(7025), 481-487. doi:10.1038/nature03235

Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M., & O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics, 196*(4), 961-971. doi:10.1534/genetics.113.160713

Groth, A. C., Fish, M., Nusse, R., & Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics, 166*(4), 1775-1782.

Horn, C., & Wimmer, E. A. (2000). A versatile vector set for animal transgenesis. *Dev Genes Evol, 210*(12), 630-637.

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., & Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell, 132*(5), 783-793. doi:10.1016/j.cell.2008.01.014

Jeong, S., Rokas, A., & Carroll, S. B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell, 125*(7), 1387-1399. doi:10.1016/j.cell.2006.04.043

John, A. V., Sramkoski, L. L., Walker, E. A., Cooley, A. M., & Wittkopp, P. J. (2016). Sensitivity of Allelic Divergence to Genomic Position: Lessons from the Drosophila tan Gene. *G3 (Bethesda), 6*(9), 2955-2962. doi:10.1534/g3.116.032029

Kopp, A. (2009). Metamodels and phylogenetic replication: a systematic approach to the evolution of developmental pathways. *Evolution, 63*(11), 2771-2789. doi:10.1111/j.1558-5646.2009.00761.x

Kopp, A., Duncan, I., Godt, D., & Carroll, S. B. (2000). Genetic control and evolution of sexually dimorphic characters in Drosophila. *Nature, 408*(6812), 553-559. doi:10.1038/35046017

Massey, J. H., & Wittkopp, P. J. (2016). The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species. *Curr Top Dev Biol, 119*, 27-61. doi:10.1016/bs.ctdb.2016.03.004

Morales-Hojas, R., Vieira, C. P., & 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. *Molecular Ecology, 17*(12), 2910-2926. doi:10.1111/j.1365-294X.2008.03796.x

Port, F., Chen, H. M., Lee, T., & 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*(29), E2967-2976. doi:10.1073/pnas.1405500111

Rasband, W. S. (1997-2016). ImageJ. Bethesda, Maryland, USA: U. S. National Institutes of Health. Retrieved from http://imagej.nih.gov/ij/

RCoreTeam. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Takahashi, A. (2013). Pigmentation and behavior: potential association through pleiotropic genes in Drosophila. *Genes Genet Syst, 88*(3), 165-174.

True, J. (2003). Insect melanism: the molecules matter. *Trends Ecol Evol, 18*(12), 640-647.

Warming, S., Costantino, N., Court, D. L., Jenkins, N. A., & Copeland, N. G. (2005). Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res, 33*(4), e36. doi:10.1093/nar/gni035

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell, 134*(4), 610-623. doi:10.1016/j.cell.2008.06.052

Wittkopp, P. J., Carroll, S. B., & Kopp, A. (2003). Evolution in black and white: genetic control of pigment patterns in Drosophila. *Trends Genet, 19*(9), 495-504. doi:10.1016/S0168-9525(03)00194-X

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., . . . Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science, 326*(5952), 540-544. doi:10.1126/science.1176980

Wittkopp, P. J., True, J. R., & Carroll, S. B. (2002). Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns. *Development, 129*(8), 1849-1858.

Wittkopp, P. J., Vaccaro, K., & Carroll, S. B. (2002). Evolution of yellow gene regulation and pigmentation in Drosophila. *Curr Biol, 12*(18), 1547-1556.

Chapter 4

Probing the mechanism through which noncoding sequences in *tan* impact pigmentation divergence between *D. americana* and *D. novamexicana*

Abstract

While changes in noncoding DNA are known to contribute to phenotypic evolution, the mechanisms through which they act aren't well understood. Increasing the number of cases in which specific genetic changes in noncoding DNA sequence are linked to specific phenotypic consequences, as well as the molecular mechanism connecting the two, will provide greater understanding into the contribution of noncoding DNA sequences to phenotypic evolution. Here, I explore potential mechanisms through which noncoding DNA sequences in *tan* may contribute to the pigmentation divergence between *D. americana* and *D. novamexicana*. Specifically, I test the hypothesis that divergent noncoding sequences in *tan* intron 1 influence pigmentation divergence by altering transcriptional level gene expression. Attempts to measure the effect of noncoding sequence changes on *tan* expression in *D. melanogaster* transgenics proved unsuccessful in detecting biologically relevant differences, perhaps due to the divergent genomic context in which the *D. americana* and *D. novamexicana tan* alleles were tested. Noncoding sequences changes in *tan* intron 1 were investigated *in silico* for predicted changes in transcription factor binding sites between *D. americana* and *D. novamexicana*. Three such binding sites overlap a derived nucleotide in *D. novamexicana* sequence; these binding sites and

their associated transcription factors (*Abd-A*, *Abd-B*, *vvl*) represent a reasonable molecular mechanism through which changes in intron 1 could contribute to transcription level gene expression and/or phenotypic output of *tan*. Lastly, green fluorescent protein (GFP) reporter constructs tested noncoding sequences for enhancer activity. This experiment suggests that *tan* intron 1 and *tan* intron 3 from both *D. americana* and *D. novamexicana* have enhancer activity. Future study in this system can specifically test for differential binding of the predicted transcription factors and potential effects on mRNA expression, as well as compare activity of the GFP reporter constructs, thereby providing information regarding the molecular mechanism through which noncoding sequences in *tan* contribute to phenotypic divergence between *D. americana* and *D. novamexicana.*

Introduction

The development of novel phenotypes results from underlying genetic changes. These mutations can occur in either coding sequences (those which encode the amino acid sequence of a protein) or noncoding sequences (those which do not encode protein sequence). The genetic code allows for inferences regarding the effect of a mutation in coding sequences of DNA to be made; mutations in coding sequences can be classified as synonymous (those which do not impact amino acid sequence), nonsynonymous (those which change an amino acid in the protein), frameshift (those which delete or add bases not in a multiple of three, thus disrupting the normal reading frame), or nonsense (those which introduce a premature stop codon, thus resulting in a shorten protein sequence). Based on the type of mutation and location in the resulting protein, a mutation in coding sequence can be hypothesized to either have an effect or not have an effect on gene function. The same, however, cannot be said for noncoding sequences of DNA.

Noncoding sequences of DNA can play a role in gene expression, influence splicing, and encode for functional RNA molecules such as transfer RNA, ribosomal RNA, and regulatory RNAs, among other functions. In studying phenotypic evolution, the most well studied of these functions is regulating gene expression. In transcriptional level gene regulation, noncoding DNA sequences can serve as *cis*-regulatory elements that control when, where, and to what extent a gene is turned on or off. If the timing, location, and/or level of expression is important to a gene's phenotypic output, changes in these *cis*-regulatory elements could contribute to phenotypic evolution. Experimental evidence confirms this hypothesis, with phenotypic divergence being associated with *cis*-regulatory divergence in multiple species (Ahmed-Braimah & Sweigart, 2015; Bastide et al., 2013; Bickel, Kopp, & Nuzhdin, 2011; Dembeck et al., 2015; Endler, Betancourt, Nolte, & Schlotterer, 2016; Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Jeong et al., 2008; Johnson et al., 2015; Koshikawa et al., 2015; Miyagi, Akiyama, Osada, & Takahashi, 2015; Ordway, Hancuch, Johnson, Wiliams, & Rebeiz, 2014; Pool & Aquadro, 2007; Prud'homme et al., 2006; Rebeiz et al., 2009; Salomone, Rogers, Rebeiz, & Williams, 2013; Takahashi, Takahashi, Ueda, & Takano-Shimizu, 2007; Takahashi & Takano-Shimizu, 2011; Telonis-Scott, Hoffmann, & Sgro, 2011; Williams et al., 2008; Wittkopp et al., 2009; Yassin et al., 2016).

Previous work (Wittkopp et al., 2009) and the transgenic studies presented in Chapter 3 clearly implicate *tan* in the pigmentation divergence between *D. americana* and *D. novamexicana*. More specifically, noncoding changes in *tan* have been shown to influence pigmentation differences between these species. The mechanism through which changes in *tan* impact pigmentation, however, is not fully understood. *tan* is differentially expressed at the mRNA level between *D. americana* and *D. novamexicana* in a manner consistent with gene function in the *Drosophila* melanin pigmentation pathway (Figure 4.1A) and species pigmentation (Figure 4.1B) (Cooley, Shefner, McLaughlin, Stewart, & Wittkopp, 2012); more specifically, *tan* contributes to dark pigmentation synthesis and is expressed more highly in the darkly pigmented *D. americana*. By producing an F1 hybrid between the two species, the

contribution of changes in *cis*-regulatory function between the two species alleles can be measured. In the F1 hybrid, *trans*-regulatory factors from both species are present, thus any difference in allele specific expression can be attributed to divergent *cis*-regulatory function. This analysis showed that a small, yet significant, part of the overall difference in *tan* expression can be attributed to divergent *cis*-regulatory function (Cooley et al., 2012). This led to the hypothesis that non-coding sequences in *tan* contribute to pigmentation divergence by altering the mRNA expression level of *tan*. I have used three different approaches to gain insight on this hypothesis: 1) Testing for differential *tan* mRNA expression in *tan* transgenic *D. melanogaster*, 2) Predicting the impact of sequence changes between species on transcription factor binding sites (TFBS), and 3) Testing noncoding sequences in *tan* from *D. americana* and *D. novamexicana* for differences in capacity to activate gene expression, with green fluorescent protein (GFP) reporter genes.

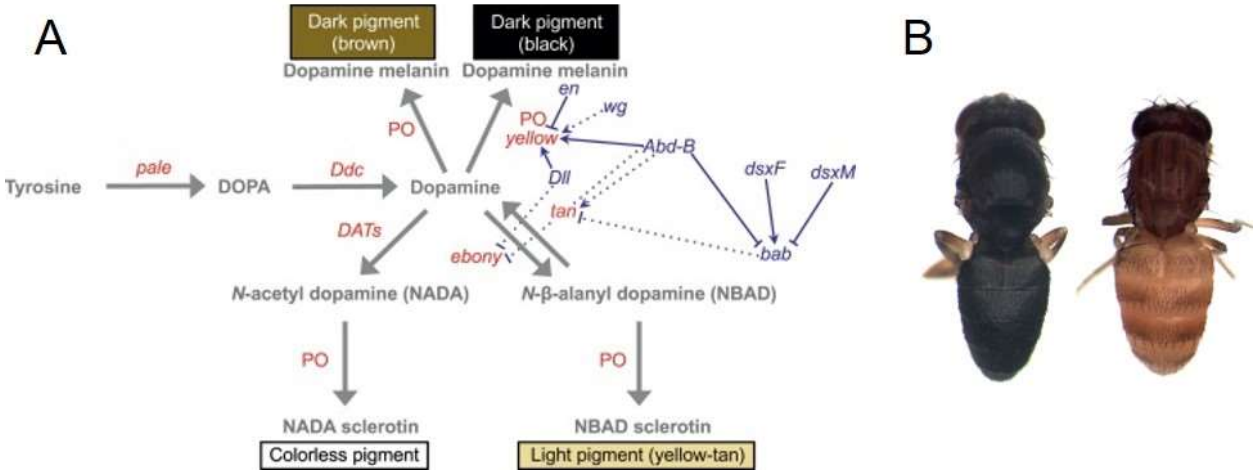

Figure 4.1: Pigmentation biosynthesis and phenotypic differences between *D. americana* and *D. novamexicana.* (A) Pigmentation biosynthesis in *Drosophila.* From Massey and Wittkopp, 2016. Genes that are part of the pigmentation biosynthesis pathway are shown in red. Genes that are involved in the regulation of pigment development are shown in blue; both direct (solid) and indirect (dashed) regulators are connected with blue lines to the gene(s) they regulate; pointed arrow connections indicate a positive regulatory interaction (i.e. activation) and blunt connection indicate a negative regulatory interaction (i.e. repression). Metabolites are shown in grey. Grey arrows indicate direction of enzymatic reactions in the pathways. (B) Body color of *D. americana* and *D. novamexicana*. *D. novamexicana* (right) has evolved lighter body pigmentation since it diverged from the common ancestor shared with *D. americana* (left). *D. americana* has retained the darker body pigmentation shared by all other members of the *virilis* group.

Materials and Methods

Fly strains

The *tan* transgenic flies used in this analysis are a subset of those presented in Chapter 3: 1) *D. americana,* 2) *D. novamexicana,* 3) *D. americana tan* with the 5' half of intron 1 replaced with *D. novamexicana* sequence (A[N_5'_intron1]), and 4) *D. americana tan* with the intron 1 replaced with *D. novamexicana* sequence (A[N_intron1]). Detailed information regarding the construction of these lines can be found within the Materials and Methods sections of Chapter 2 and Chapter 3.

Measuring relative expression of tan *transgenes*

To test for differences in expression level of various *tan* transgenes in *D. melanogaster*, relative expression levels of each transgene were measured using pyrosequencing (Wittkopp, 2011). Specifically, I measured the mRNA abundance of the *D. americana* allele relative to the mRNA abundance of the *D. novamexicana tan* allele using a divergent nucleotide between these two alleles.

For each *tan* transgenic line, both genomic DNA and total RNA were extracted from three replicate samples, each containing 12 dissected F1 hybrid pupae. Creating a F1 hybrid brings the *D. americana* and *D. novamexicana* alleles into the same *trans*-regulatory background, thus allowing the *cis*-regulatory contribution to expression differences to be tested. The P14-P15 pupal stage was analyzed because *D. americana* and *D. novamexicana tan* have previously been shown to be most highly expressed during this time (Cooley et al., 2012). Pupal heads and wings were removed to avoid measuring *tan* expression in those tissues, focusing our

measurements on expression in the thorax and abdomen where pigmentation phenotypes are most apparent.

cDNA was reverse transcribed from extracted RNA using a polyT primer for each sample. Both genomic DNA and cDNA were analyzed by pyrosequencing as described in Wittkopp et al. (Wittkopp, 2011). Expression was measured in two different exons, exon 7 and exon 8. PCR primers used to amplify the sequences analyzed are listed in Table 4.1. For the exon 7 assay, PCR primer 2 was biotinylated. In the exon 8 assay, a universal pyrosequencing primer strategy was employed (Aydin, Toliat, Bahring, Becker, & Nurnberg, 2006; Guo & Milewicz, 2007; Pacey-Miller & Henry, 2003; Royo, Hidalgo, & Ruiz, 2007). In this case, PCR primer 1 contains a tail (underlined sequence in Table 2) complementary to a biotinylated universal pyrosequencing primer. In the PCR amplification for the exon 8 assay, PCR primer 1, PCR primer 2, and the biotinylated universal pyrosequencing primer were included.

Table 4.1: PCR amplification primers and pyrosequencing primers used to measure relative expression.

Location	**PCR primer 1**	**PCR primer 2**	**Pyrosequencing primer**
Exon 7	GATGCTGAAGTCCAGCGTGTC	CAGCCGCCAGTGACATCA	CGAGCACGATGTCCG
Exon 8	GTGACGTACTAGCAACGGATGCTGAAGTCCAGCGTGTC	AGGGGCAACGTGCAGTGT	CAGGCCAACAGCAAT

Universal forward primer: [Btn]5'-GTGACGTACTAGCAACG-3'

Analysis of pyrosequencing data

For each assay, the mRNA abundance ratio of *D. americana* to *D. novamexicana* was calculated by dividing the peakheight corresponding to the incorportation at the *D. americana* nucleotide by the peakheight corresponding to the incorportation *D. novamexicana* nucleotide. In the gDNA samples from F1 hybrids, this value is expected to be equal to 1, however, slight deviations from this can occur due to PCR bias towards a particular allele. To correct for this, each cDNA ratio

was divided by its corresponding gDNA ratio to obtained a corrected measure of ratio of allelic abundance. The average, standard deviation, 95% confidence interval, and p-value (two tailed t-test assuming different variance) were calculated and are shown in Table 4.2. The $\log_2$ ratio of mRNA abundance and 95% confidence intervals were also calculated to obtain the graphical representation of the data presented in the Results and Discussion.

Table 4.2: Relative expression of *D. americana* and *D. novamexicana tan* as measured by pyrosequencing. cDNA/gDNA is the ratio of *D. americana* to *D. novamexicana* alleles measured in the sample. For this measurement, a value equal to one indicates equal expression of the two alleles. A cDNA/gDNA less than 1 indicates higher expression of the *D. novamexicana tan* allele.

Exon 7	**T=amer**	**C=nova**	**amer/nova**	**cDNA/gDNA**	**Average**	**StDev**	**95% CI**	**p value**
cDNA-1	17.98	17.35	1.04	0.88	0.88	0.01	0.01	0.0004
gDNA-1	28.97	24.48	1.18					
cDNA-2	14.83	13.78	1.08	0.89				
gDNA-2	22.41	18.49	1.21					
cDNA-3	23.2	22.51	1.03	0.88				
gDNA-3	19.43	16.53	1.18					

Exon8	**T=amer**	**C=nova**	**amer/nova**	**cDNA/gDNA**	**Average**	**SD**	**95% CI**	**p value**
cDNA-1	6.32	9.44	0.67	0.68	0.76	0.06	0.07	0.0074
gDNA-1	11.13	11.29	0.99					
cDNA-2	3.91	4.75	0.82	0.81				
gDNA-2	9.85	9.74	1.01					
cDNA-3	8.88	10.33	0.86	0.78				
gDNA-3	11.22	10.22	1.10					

Prediction of changes in transcription factor binding site

A list of 28 candidate transcription factors was curated from Rogers et al. (2013), which describes a RNAi screen of transcription factors in *D. melanogaster* that caused reduced or ectopic abdominal pigmentation. Of these 28 transcription factors with evidence for involvement in abdominal pigmentation, 17 have predicted transcription factor binding motifs (Shazman et al., 2014). With results presented in Chapter 3 suggesting the 5' half of intron 1 of *tan* contains sequence contributing to pigmentation divergence between *D. americana* and *D. novamexicana*, this DNA sequence was searched for predicted transcription factor binding sites (TFBSs) for TFs implicated in Rogers et al. (2013). Differences in transcription factor binding sites that

overlapped fixed sequence differences between species were identified. Information regarding inferred effect on pigmentation, whether the difference between species predicts strengthening or weakening of transcription factor binding, and whether the fixed sequence difference is derived in *D. americana* or *D. novamexicana* was used to prioritize transcription factor binding sites. Additionally, a 20bp sequence centered on the top candidate SNP in the 5' half of *tan* intron 1 was input as a query motif in Tomtom (http://meme-suite.org/tools/tomtom) to generate alignments of this sequence to known transcription factor binding site information in *D. melanogaster*.

Generation of GFP reporter transgenic flies

For both *D. americana* and *D. novamexicana*, intron 1 and intron 3 were cloned into a previously constructed GFP reporter plasmid (Kalay & Wittkopp, 2010) containing an enhanced green fluorescent protein (EGFP) under control of the *hsp70* promoter, an attB sequenced used for ΦC31-mediated transformation into the *D. melanogaster* genome, and the eye-expressing 3xP3-EGFP marker for detection of successful integration. Specifically, primers that amplify intron 1 and intron 3 were created with 20bp homology tails flanking the unique XbaI site in the plasmid. After PCR amplification with these primers, each fragment was inserted into the GFP reporter plasmid via Gibson Assembly (Gibson, 2011; Gibson et al., 2009). Fully constructed reporter constructs were sequence confirmed using Sanger sequencing performed by the University of Michigan Sequencing Core (https://seqcore.brcf.med.umich.edu/).

GenetiVision (Houston, Texas) performed all DNA preparations and embryo injections into a *D. melanogaster* host according to their standard protocols (http://www.genetivision.com/). The *D. melanogaster* host genotype carried a transgene on the

X-chromosome using the *vasa cis*-regulatory sequences to express the ΦC31 integrase specifically in the germ-line, a single attP site located on the third (86Fb – BDSC #24749) chromosome, and an eye-expressing red fluorescent protein (RFP) as a visible marker for the landing site (Bischof, Maeda, Hediger, Karch, & Basler, 2007). Injected males were mated to virgin *yellow, white, tan* mutant females and resulting male offspring were screened for the transformation marker (green fluorescent protein expressed in their eyes). This cross prior to screening was required to remove the GFP-marker *vasa* driven ΦC31 integrase. Transformant males were used to establish lines homozygous for each transgene in a *D. melanogaster* background carrying loss-of-function mutations in the X-linked genes *tan, yellow,* and *white*.

Analysis of GFP reporter gene expression patterns

Appropriately staged (P14) pupae were identified by the visible pigmentation developing the wings and the location of the meconium in the abdomen (Bainbridge & Bownes, 1981). Pupal bodies were prepared for confocal microscopy by removing the pupae from their pupal casing while taking care to prevent damaging the transparent pupal cuticle surrounding each pupa. Each pupa was mounted on microscope slide in a drop of water and covered with a coverslip. Pupae were imaged within 1 hour of mounting using a Leica Sp5 confocal microscope using an argon laser to at 25% power to detect GFP. Maximum projections from the z-stack projection were saved and processed identically.

Results and Discussion

Testing for differential tan mRNA expression using transgenic D. melanogaster

To test the hypothesis that noncoding sequences changes contribute to pigmentation divergence between *D. americana* and *D. novamexicana* by affecting *cis*-regulatory function, I measured the

mRNA expression level of *tan* using *tan* transgenic *D. melanogaster* lines and tested for differential expression. The *tan* alleles used in this study are: *D. americana, D. novamexicana, D. americana tan* with the 5' half of intron 1 replaced with *D. novamexicana* sequence (A[N_5'_intron1]), and *D. americana tan* with the intron 1 replaced with *D. novamexicana* sequence (A[N_intron1]). With these *tan* chimeric alleles that change only intron 1, the previously fine-mapped region implicated in pigmentation divergence between species (Wittkopp et al., 2009), I can directly test the impact of these sequences on *tan* expression level and gain insights into the mechanism these noncoding changes are working through.

Using pyrosequencing, I quantified the relative mRNA expression level between *tan* transgenic lines during the late pupal stages when pigmentation is developing (Cooley et al., 2012). As a pilot experiment, only the *D. americana* and *D. novamexicana tan* transgenic lines were used. If a difference in *tan* expression can be detected between these lines, then more subtle differences driven by changes in noncoding sequences in the chimeric *tan* alleles may be able to be distinguished. However, if I am unable to measure significant differences between the *D. americana* and *D. novamexicana tan* transgenic lines our attempts to detect differences driven by the chimeric *tan* alleles is unlikely to be successful.

D. americana and *D. novamexicana tan* transgenic lines of *D. melanogaster* were crossed to produce an F1 hybrid population. Any differential expression between *tan* alleles detected in these flies can be attributed to changes in *cis*-regulatory function. Given that differences in pigmentation are seen in these transgenic constructs and that a *cis*-regulatory contribution, albeit a small one, to gene expression difference measured between species (Cooley et al., 2012), it is hypothesized that a significant difference in *tan* expression will be detected. More specifically, it

is hypothesized that the *D. americana* allele of *tan* will be expressed more highly than the *D. novamexicana tan* allele.

Despite our ability to detect pigmentation differences between the *D. americana* and *D. novamexicana tan* transgenic *D. melanogaster* lines, experiments to test for differential *tan* mRNA expression failed to support our hypothesis. Instead, *tan* expression measured in two different exons both showed the *D. novamexicana tan* allele being expressed more highly than the *D. americana tan* allele (Figure 4.2). This difference was significantly different from a null expectation of equal expression in both cases (Exon 7, p=0.0004; Exon 8, p = 0.007) using a t-test assuming unequally variances.

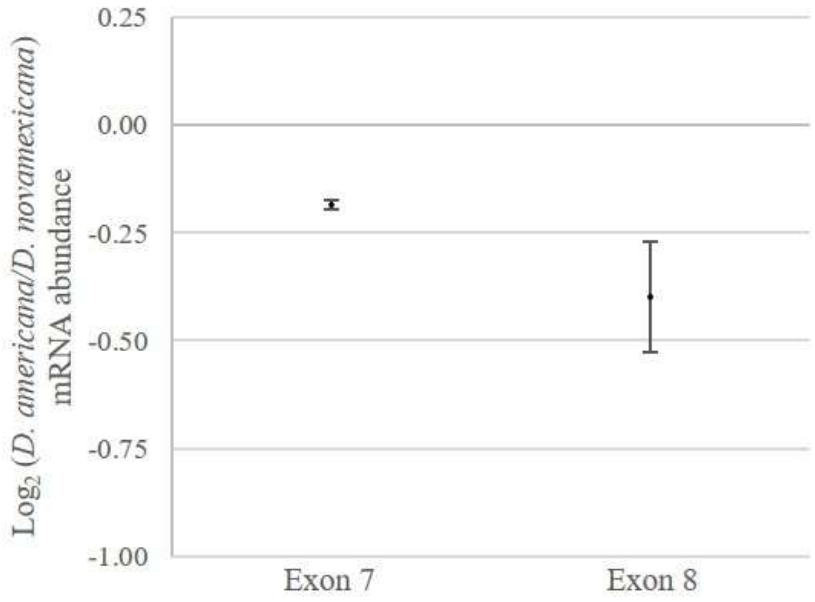

Figure 4.2: Relative expression of *D. americana* and *D. novamexicana tan* alleles. Log2 values of the *D. americana* to *D. novamexicana tan* alleles are plotted as measured in Exon 7 and Exon 8. Values less than 0 indicate *D. novamexicana* is expressed more highly than *D. americana*. Error bars represent the 95% confidence interval.

These results suggest that the difference in pigmentation driven by *D. americana* and *D. novamexicana tan* transgenes in *D. melanogaster* does not act at the level of mRNA expression. This result is surprising when previous work showing that the *D. americana tan* allele had a

modest (29%), but significant, increase in *cis*-regulatory function compared to *D. novamexicana* in their native species (Cooley et al., 2012) is taken into account. Given that the virilis group (which includes *D. americana* and *D. novamexicana*) and *D. melanogaster* last shared a common ancestor ~40 million years ago, it seems reasonable that the *trans*-regulatory factors background between these species have amassed significant differences. Exploring the molecular mechanism through which noncoding sequence changes impart their effect on pigmentation in such a distinct genetic background may not accurately reflect the underlying biology in the native species. With the advent of CRISPR technology, creating specific genetic changes in non-model species has become a reality (Huang, Liu, & Rong, 2016). Future study in which these non-coding sequence changes are create in the *D. americana* genomic background may provide more accurate description of the molecular mechanism at play and the specific effects of noncoding sequences of *tan* on mRNA-level gene expression.

In silico testing for predicted changes in transcription factor binding sites (TFBSs) between species

If the main hypothesis that non-coding sequences in *tan* contribute to pigmentation divergence by altering the mRNA expression level of *tan* was supported, I hypothesized that changes in transcription factor binding sites (TFBSs) within intron 1 would exist between *D. americana* and *D. novamexicana.* Using a previously curated a list of transcription factors (TFs) that appear to be involved in melanin pigmentation in *D. melanogaster* as suggested by RNAi knockdown of the TF causing a change in pigmentation (Rogers et al., 2014), I surveyed intron 1 sequence from *D. americana* and *D. novamexicana* for predicted binding sites for this set of TFs (Shazman, Lee, Socol, Mann, & Honig, 2014). Difference in these predicted TFBSs between species were

identified and prioritized for potential influence on mRNA expression using information regarding TF influence on pigmentation (i.e. does RNAi knock-down increase or decrease pigmentation), as well as evolutionary relationships to determine which sequence changes were derived in *D. novamexicana*, the species with the derived phenotype. Specifically, I utilized *D. virilis* sequence to determine which alleles were derived in *D. americana* and which were derived in *D. novamexicana*. For instance, if *D. americana* and *D. virilis* have the same nucleotide at a particular position, but *D. novamexicana* has a different nucleotide, that site is said to be derived in *D. novamexicana*.

Analysis focused on the 5' half of intron 1 for this analysis due to the impact this sequence had on abdominal pigmentation in the *D. melanogaster tan* transgenics. The most likely candidates for nucleotides contributing to pigmentation divergence between *D. americana* and *D. novamexicana* are those derived in *D. novamexicana*, the species with the derived phenotype. In the 5' half of intron 1, four fixed sequence differences exist that are derived changes in *D. novamexicana*. By utilizing additional sequence information from *D. americana*, these changes can be further prioritized for examining potential changes in TFBSs. A *D. americana* line isolated from the wild exists that exhibits a light pigmentation phenotype relative to majority of *D. americana* lines. In previous studies, this line was shown to have a *tan* allele that was functionally equivalent to the *D. novamexicana* allele of *tan* (Wittkopp et al., 2009). Therefore, it is expected that *D. novamexicana* and this lightly pigmented *D. americana* line to share the causative nucleotide(s). When taking this information into account, a single sequence change emerges as the highest priority candidate.

Using Tomtom (http://meme-suite.org/tools/tomtom), a TFBS motif comparison tool, the region surrounding this nucleotide change was probed for possible TFBSs and differences in

these predictions between *D. americana* and *D. novamexicana*. At the candidate nucleotide, there are three predicted TFBSs that either have reduced affinity or are abolished completely based on the predicted sequence motifs. The three transcription factors that are predicted to have binding sites around this nucleotide are: *AbdA, AbdB,* and *vvl.* Each of these has a predicted binding site in *D. americana* that is abolished (*AbdB, vvl*) or predicted reduced affinity (*AdbA*) in *D. novamexicana* (Figure 4.3). Additionally, when the expression of these transcription factors is knocked down by RNAi, abdominal pigmentation decreases suggesting that these transcription factors are important for dark pigment formation in the developing fly (Rogers et al., 2014). Furthermore, *AbdA* and *AbdB* often work with a cofactor *exd* and when they do, their binding sites are often very close or overlapping with one another (Camino et al., 2015; Mann, Lelli, & Joshi, 2009; Slattery et al., 2011). A predicted binding site for *exd* exists in the sequence directly adjacent (and partially overlapping) with the predicted binding sites for *AbdA* and *AbdB*, giving further support for the hypothesis that these transcription factors may be regulating gene expression by binding to this sequence in *tan* intron 1.

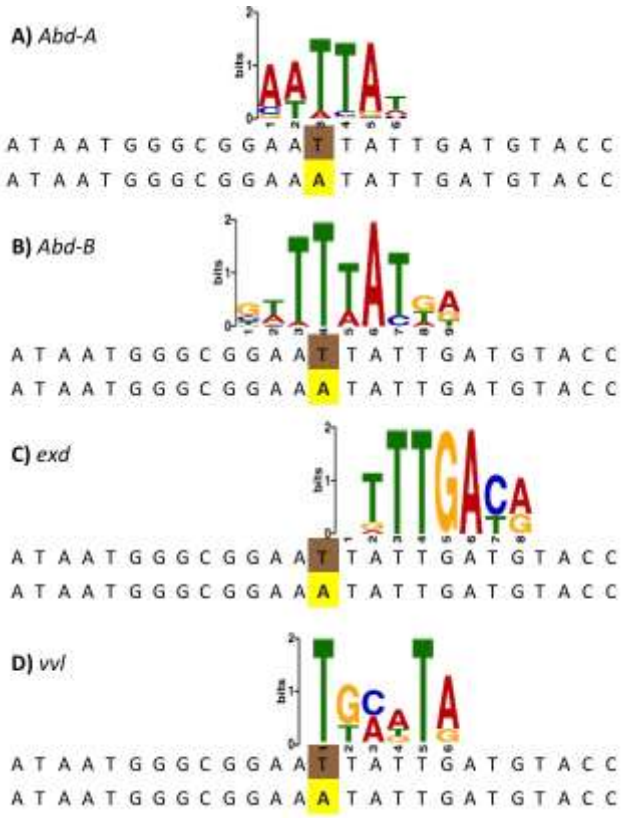

Figure 4.3: Predicted transcription factor binding sites (TFBSs) surrounding candidate SNP in *tan* intron 1. The predicted TFBS motif is shown above the *D. americana* (brown) and *D. novamexicana* (yellow) sequences.

This nucleotide and the changes in predicted TFBSs accompanying it represent the highest priority candidates for a molecular mechanism by which a noncoding change in intron 1 of *tan* can impact pigmentation via altered *cis*-regulatory function between *D. americana* and *D. novamexicana*. Future studies should test both the nucleotide and the identified potential transcription factors to determine their functional effect (if any) on pigmentation. These studies could include: 1) performing assays to test for differential binding of the transcription factors to the sequence flanking the nucleotide, and 2) using CRISPR to edit the nucleotide in *D. americana* and measuring the pigmentation effect. Combined, these studies would provide valuable information regarding if this nucleotide impacts TF binding of any of its predicted regulators and if changing this nucleotide impacts pigmentation differences between *D.*

americana and *D. novamexicana*. Based on the results of these experiments, additional experiments could be designed to test the effect of this nucleotide on the mRNA expression level of *tan*.

GFP reporter genes test for functional differences between noncoding sequences in D. americana and D. novamexicana tan

While testing the phenotypic impact of noncoding sequence changes in *tan* definitively links those sequences to pigmentation divergence between *D. americana* and *D. novamexicana*, other experiments can provide information regarding if there are functional differences between these noncoding sequences between species. I hypothesized that non-coding sequences in *tan* contribute to pigmentation divergence by altering the mRNA expression level of *tan*. Underlying this is an additional hypothesis that *tan* intron 1 contains an enhancer. To test this hypothesis, reporter constructs in which *D. americana* or *D. novamexicana* noncoding sequence is used to drive expression of green fluorescent protein (GFP) were constructed. These reporter constructs will allow us to test for the presence of enhancers in these noncoding sequences as well as test for differences in the ability of *D. americana* and *D. novamexicana* sequence to drive GFP expression. As such, I have created GFP reporter constructs which use either *D. americana* or *D. novamexicana tan* intron 1 to drive GFP expression in *D. melanogaster*. In *tan*, intron 1 and 3 are relatively large compared to the other introns and therefore seem more likely to contain *cis*-regulatory elements. This, in addition to previous work suggesting that other noncoding sequences in *tan* contribute to the pigmentation divergence between species, motivated my decision to construct additional GFP reporters using either *D. americana* or *D. novamexicana tan* intron 3 to drive GFP expression.

Preliminary results from examination of these GFP reporter constructs indicate that each of the noncoding regions tested have enhancer function since they drove faint, but detectable, GFP expression compared to the no enhancer control (Figure 4.4). However, conclusion regarding relative strength of these enhancers between species have not been made yet. At the time of imaging, the flies containing the GFP reporter construct were not yet homozygous and instead contained a mix of homozygous and heterozygous individuals. As such, some of the individuals imaged may have two copies of the reporter construct while other only have one, making direct comparisons between constructs uninformative. Once homozygous, the GFP expression pattern and intensity from these lines will be imaged, quantified, and systematically compared.

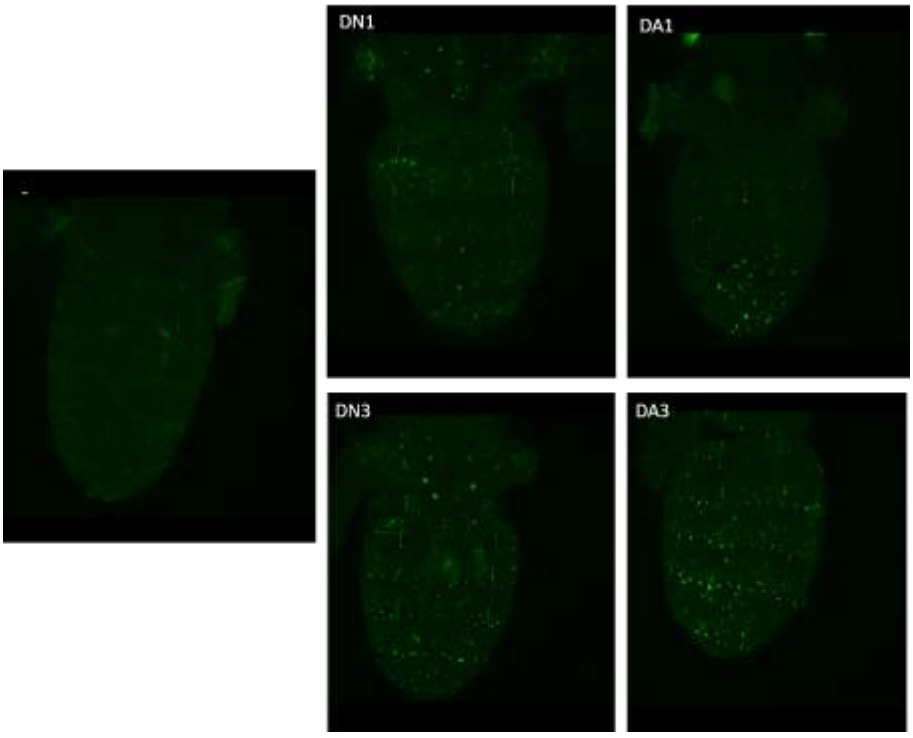

Figure 4.4: Green fluorescent protein (GFP) expression patterns driven by noncoding sequences in *tan*. (-) is a no enhancer control. DN1 uses *D. novamexicana tan* intron 1 to drive GFP expression. DA1 uses *D. americana tan* intron 1 to drive GFP expression. DN3 uses *D. novamexicana tan* intron 3 to drive GFP expression. DA3 uses *D. americana tan* intron 3 to drive GFP expression.

Acknowledgments

This work would not have been possible without: Lisa Sramkoski and Arielle Cooley for their contributions to the foundation of this work, including generation of *tan* transgenics; Bradley Lankowsky, Wesley McLaughlin, Robert Dikeman, Elizabeth Walker, and Abigail Lamb for assistance in transgenic construction, pupal dissections, and general fly husbandry; Abigail Lamb for experimental assistance with dissections, DNA/RNA extractions, pyrosequencing, and transcription factor binding analysis; Jennifer Lachowiec for assistance with pupal dissections and confocal microscopy; and members of the Dissertation Committee including Scott Barolo, Gyorgyi Csankovszki, Monica Dus, and Andrzej Wierzbicki (former dissertation committee member) as well as the Wittkopp lab for helpful discussions and experimental design suggestions. This work was funded by the National Institute of Health (R01-GM-089736).

References

Ahmed-Braimah, Y. H., & Sweigart, A. L. (2015). A single gene causes an interspecific difference in pigmentation in Drosophila. *Genetics, 200*(1), 331-342. doi:10.1534/genetics.115.174920

Aydin, A., Toliat, M. R., Bahring, S., Becker, C., & Nurnberg, P. (2006). New universal primers facilitate Pyrosequencing. *Electrophoresis, 27*(2), 394-397. doi:10.1002/elps.200500467

Bainbridge, S. P., & Bownes, M. (1981). Staging the metamorphosis of Drosophila melanogaster. *J Embryol Exp Morphol, 66*, 57-80.

Bastide, H., Betancourt, A., Nolte, V., Tobler, R., Stobe, P., Futschik, A., & Schlotterer, C. (2013). A genome-wide, fine-scale map of natural pigmentation variation in Drosophila melanogaster. *PLoS Genet, 9*(6), e1003534. doi:10.1371/journal.pgen.1003534

Bickel, R. D., Kopp, A., & Nuzhdin, S. V. (2011). Composite effects of polymorphisms near multiple regulatory elements create a major-effect QTL. *PLoS Genet, 7*(1), e1001275. doi:10.1371/journal.pgen.1001275

Bischof, J., Maeda, R. K., Hediger, M., Karch, F., & Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A, 104*(9), 3312-3317. doi:10.1073/pnas.0611511104

Camino, E. M., Butts, J. C., Ordway, A., Vellky, J. E., Rebeiz, M., & Williams, T. M. (2015). The evolutionary origination and diversification of a dimorphic gene regulatory network through parallel innovations in cis and trans. *PLoS Genet, 11*(4), e1005136. doi:10.1371/journal.pgen.1005136

Cooley, A. M., Shefner, L., McLaughlin, W. N., Stewart, E. E., & Wittkopp, P. J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in Drosophila americana and D. novamexicana. *Evol Dev, 14*(4), 317-325. doi:10.1111/j.1525-142X.2012.00550.x

Dembeck, L. M., Huang, W., Magwire, M. M., Lawrence, F., Lyman, R. F., & Mackay, T. F. (2015). Genetic Architecture of Abdominal Pigmentation in Drosophila melanogaster. *PLoS Genet, 11*(5), e1005163. doi:10.1371/journal.pgen.1005163

Endler, L., Betancourt, A. J., Nolte, V., & Schlotterer, C. (2016). Reconciling Differences in Pool-GWAS Between Populations: A Case Study of Female Abdominal Pigmentation in Drosophila melanogaster. *Genetics, 202*(2), 843-855. doi:10.1534/genetics.115.183376

Gibson, D. G. (2011). Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol, 498*, 349-361. doi:10.1016/B978-0-12-385120-8.00015-2

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods, 6*(5), 343-345. doi:10.1038/nmeth.1318

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature, 433*(7025), 481-487. doi:10.1038/nature03235

Guo, D. C., & Milewicz, D. M. (2007). Universal primer applications for pyrosequencing. *Methods Mol Biol, 373*, 57-62. doi:10.1385/1-59745-377-3:57

Huang, Y., Liu, Z., & Rong, Y. S. (2016). Genome Editing: From Drosophila to Non-Model Insects and Beyond. *J Genet Genomics, 43*(5), 263-272. doi:10.1016/j.jgg.2016.04.007

Jeong, S., Rebeiz, M., Andolfatto, P., Werner, T., True, J., & Carroll, S. B. (2008). The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. *Cell, 132*(5), 783-793. doi:10.1016/j.cell.2008.01.014

Johnson, W. C., Ordway, A. J., Watada, M., Pruitt, J. N., Williams, T. M., & Rebeiz, M. (2015). Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between Drosophila Species. *PLoS Genet, 11*(6), e1005279. doi:10.1371/journal.pgen.1005279

Kalay, G., & Wittkopp, P. J. (2010). Nomadic enhancers: tissue-specific cis-regulatory elements of yellow have divergent genomic positions among Drosophila species. *PLoS Genet, 6*(11), e1001222. doi:10.1371/journal.pgen.1001222

Koshikawa, S., Giorgianni, M. W., Vaccaro, K., Kassner, V. A., Yoder, J. H., Werner, T., & Carroll, S. B. (2015). Gain of cis-regulatory activities underlies novel domains of wingless gene expression in Drosophila. *Proc Natl Acad Sci U S A, 112*(24), 7524-7529. doi:10.1073/pnas.1509022112

Mann, R. S., Lelli, K. M., & Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol, 88*, 63-101. doi:10.1016/S0070-2153(09)88003-4

Miyagi, R., Akiyama, N., Osada, N., & Takahashi, A. (2015). Complex patterns of cis-regulatory polymorphisms in ebony underlie standing pigmentation variation in Drosophila melanogaster. *Molecular Ecology, 24*(23), 5829-5841. doi:10.1111/mec.13432

Ordway, A. J., Hancuch, K. N., Johnson, W., Wiliams, T. M., & 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, 392*(2), 431-440. doi:10.1016/j.ydbio.2014.05.023

Pacey-Miller, T., & Henry, R. (2003). Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. *Anal Biochem, 317*(2), 166-170.

Pool, J. E., & Aquadro, C. F. (2007). The genetic basis of adaptive pigmentation variation in Drosophila melanogaster. *Molecular Ecology, 16*(14), 2844-2851. doi:10.1111/j.1365-294X.2007.03324.x

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., . . . Carroll, S. B. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature, 440*(7087), 1050-1053. doi:10.1038/nature04597

Rebeiz, M., Ramos-Womack, M., Jeong, S., Andolfatto, P., Werner, T., True, J., . . . Carroll, S. B. (2009). Evolution of the tan locus contributed to pigment loss in Drosophila santomea: a response to Matute et al. *Cell, 139*(6), 1189-1196. doi:10.1016/j.cell.2009.11.004

Rogers, W. A., Grover, S., Stringer, S. J., Parks, J., Rebeiz, M., & Williams, T. M. (2014). A survey of the trans-regulatory landscape for Drosophila melanogaster abdominal pigmentation. *Dev Biol, 385*(2), 417-432. doi:10.1016/j.ydbio.2013.11.013

Royo, J. L., Hidalgo, M., & Ruiz, A. (2007). Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. *Nat Protoc, 2*(7), 1734-1739. doi:10.1038/nprot.2007.244

Salomone, J. R., Rogers, W. A., Rebeiz, M., & Williams, T. M. (2013). The evolution of Bab paralog expression and abdominal pigmentation among Sophophora fruit fly species. *Evol Dev, 15*(6), 442-457. doi:10.1111/ede.12053

Shazman, S., Lee, H., Socol, Y., Mann, R. S., & Honig, B. (2014). OnTheFly: a database of Drosophila melanogaster transcription factors and their binding sites. *Nucleic Acids Res, 42*(Database issue), D167-171. doi:10.1093/nar/gkt1165

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., . . . Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell, 147*(6), 1270-1282. doi:10.1016/j.cell.2011.10.053

Takahashi, A., Takahashi, K., Ueda, R., & Takano-Shimizu, T. (2007). Natural variation of ebony gene controlling thoracic pigmentation in Drosophila melanogaster. *Genetics, 177*(2), 1233-1237. doi:10.1534/genetics.107.075283

Takahashi, A., & Takano-Shimizu, T. (2011). Divergent enhancer haplotype of ebony on inversion In(3R)Payne associated with pigmentation variation in a tropical population of Drosophila melanogaster. *Molecular Ecology, 20*(20), 4277-4287. doi:10.1111/j.1365-294X.2011.05260.x

Telonis-Scott, M., Hoffmann, A. A., & Sgro, C. M. (2011). The molecular genetics of clinal variation: a case study of ebony and thoracic trident pigmentation in Drosophila melanogaster from eastern Australia. *Molecular Ecology, 20*(10), 2100-2110. doi:10.1111/j.1365-294X.2011.05089.x

Williams, T. M., Selegue, J. E., Werner, T., Gompel, N., Kopp, A., & Carroll, S. B. (2008). The regulation and evolution of a genetic switch controlling sexually dimorphic traits in Drosophila. *Cell, 134*(4), 610-623. doi:10.1016/j.cell.2008.06.052

Wittkopp, P. J. (2011). Using pyrosequencing to measure allele-specific mRNA abundance and infer the effects of cis- and trans-regulatory differences. *Methods Mol Biol, 772*, 297-317. doi:10.1007/978-1-61779-228-1_18

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., . . . Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science, 326*(5952), 540-544. doi:10.1126/science.1176980

Yassin, A., Bastide, H., Chung, H., Veuille, M., David, J. R., & Pool, J. E. (2016). Ancient balancing selection at tan underlies female colour dimorphism in Drosophila erecta. *Nat Commun, 7*, 10400. doi:10.1038/ncomms10400

Chapter 5

Determining the genetic basis of pigmentation diversity within *Drosophila americana*

Abstract

In evolutionary biology, many unanswered questions about similarities and differences in phenotypic evolution within and between species exist. Ideal systems in which to study these similarities and differences would be those in which interspecies variation and intraspecies variation in the same phenotype exist. One such system is comparing the interspecies pigmentation divergence between *Drosophila americana* and *D. novamexicana* with the pigmentation diversity within *D. americana.* Here, the genetic basis of intraspecific pigmentation diversity in *D. americana* is explored through a candidate gene approach. All pairwise combinations of two light and two dark strains of *D. americana* were used to generate F2 hybrid populations. F2 hybrid individuals were sorted into pigmentation classes and allele frequencies of the pigmentation genes *tan*, *ebony*, and *yellow* were measured in the most extreme phenotypic tails from each cross. These results suggest three main conclusions: 1) *tan* and *ebony* contribute to phenotypic diversity, while *yellow* does not, 2) other genetic loci contribute to pigmentation divergence within *D. americana*, and 3) genetic heterogeneity exists within phenotypically similar strains of *D. americana.*

Introduction

Understanding the genetic basis of phenotypic evolution is an important quest in biology. Investigating this subject can include elucidating the specific genetic changes that lead to the development of new phenotypes, discovering the molecular mechanism through which new phenotypes arise, and identifying general trends in how new phenotypes come about. In terms of identifying general trends in phenotypic evolution, an interesting question is whether the same or different factors are at play within and between species. More specifically, common questions include: 1) Are the same or different genes involved in intraspecies and interspecies phenotypic divergence? 2) If the same genes are utilized, are the same or different nucleotides responsible? 3) Are changes in noncoding or coding sequences more likely to be responsible for phenotypic divergence? 4) When changes in gene regulation contribute to phenotypic divergence, are changes in *cis*-regulatory sequences or *trans*-regulatory factors involved most often? As with other questions regarding phenotypic evolution, pigmentation in *Drosophila* is a commonly used system to explore questions regarding intraspecies phenotypic variation. When considering the evolution of pigmentation in *Drosophila*, several genes have been repeatedly utilized in pigmentation divergence within and between species, including *bric-a-brac, ebony, tan,* and *yellow* (Massey & Wittkopp, 2016). However, the vast majority of study within a single species has taken place in *D. melanogaster*, where *bric-a-brac*, *ebony*, and *tan* have been repeatedly implicated in intraspecies pigmentation variation. It is unclear whether this is a general trend *Drosophila* pigmentation evolution or a result of the emphasis of study in *D. melanogaster*. Study of additional species exhibiting intraspecies pigmentation variation would increase our understanding of the similarities and differences in phenotypic divergence within and between species.

In *D. americana*, pigmentation varies significantly with longitude; populations of *D. americana* exhibit lighter pigmentation as you move west across the continental United States((Figure 5.1) (Wittkopp et al., 2011). This type of geographic variation could be caused by either neutral (i.e. genetic drift) or non-neutral (i.e. natural selection) evolutionary processes. Previous work found that the association between pigmentation and geographic location in *D. americana* was inconsistent with genetic drift; instead, the pigmentation cline likely results from natural selection favoring particular genotypes in specific geographic regions despite gene flow between *D. americana* populations acting to homogenize the genome (Wittkopp et al., 2011). When taken with the information already known about the pigmentation divergence between *D. americana* and *D. novamexicana*, this system is ideal for exploring questions regarding intra- and inter-specific phenotypic divergence.

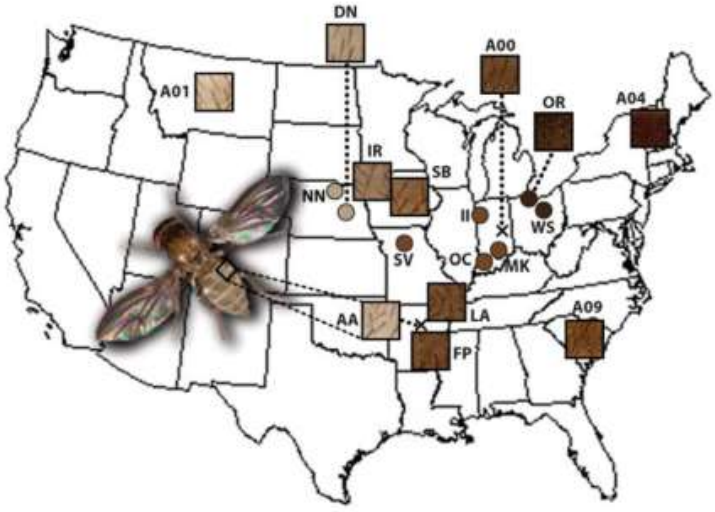

Figure 5.1 Pigmentation in *D. americana* varies with geographical location. From Wittkopp et al. 2010. A pigmentation cline within *D. americana* exists such that populations become increasingly lighter in pigmentation moving west across the continental United States.

In *D. americana* and *D. novamexicana*, the combined effects of *tan* and *ebony* are responsible for approximately 87% of the pigmentation divergence between species (Wittkopp et al., 2009). During speciation, it is likely that natural selection acted on this existing genetic variation in *tan* and *ebony* leading to the fixed differences in functionality seen in these genes between *D. americana* and *D. novamexicana.* Other pigmentation genes, including *yellow* and *bric-a-brac*, were not found to contribute to pigmentation divergence between *D. americana* and *D. novamexicana* (Wittkopp et al., 2009). As in the pigmentation divergence between *D. americana* and *D. novamexicana*, the pigmentation genes *tan* and *ebony* likely contribute to the intraspecific pigmentation diversity within *D. americana* (Wittkopp et al., 2009). This does not, however, indicate that there aren't additional genotypes underlying the pigmentation cline within *D. americana*.

We hypothesize that *tan* and *ebony* contribute to intraspecies pigmentation variation in *D. americana*, as suggested by previous study (Wittkopp et al., 2009). Additionally, we hypothesize that genetic variation at other genes also contributes to the pigmentation cline in *D. americana* since *tan* and *ebony* do not explain the full phenotypic difference seen between *D. americana* and *D. novamexicana*; therefore, it seems logical that other genes would also contribute to the intraspecies pigmentation variation within *D. americana.* Finally, we hypothesize that genetic heterogeneity exists in *D. americana* such that independent lines isolated from the wild have different genotypes underlying similar phenotypes. Genetic heterogeneity is the biological phenomena in which similar phenotypes are produced by different genes or alleles. Support for this hypothesis is provided by a *D. americana* line isolated from Nebraska (DN2) having a *D. novamexicana*-like allele at *ebony*, but other, phenotypically similar lines isolated from the same geographic location do not share this allele (Wittkopp et al., 2009). Additional support is

provided by *D. americana* isofemale lines isolated from the same geographic location and exhibiting similar pigmentation phenotypes producing different numbers of phenotypic classes when used to create backcross populations with the lightly pigmented *D. novamexicana* (Arnold & McLaughlin, unpublished data). To create this backcross population, *D. americana* females from different strains were mated to *D. novamexicana* males; the F1 hybrid females from this cross were then backcrossed to *D. novamexicana* males to create a backcross population (BC1). The presence of differing numbers of phenotypic classes in the BC1 populations between *D. americana* strains suggests that recombination throughout the genome of the F1 hybrid brought together different combinations of alleles in different *D. americana* strains. If the same genotype was responsible for the similar phenotype in all similarly pigmented lines, we would have seen the same number of phenotypic classes for each cross (Arnold & McLaughlin, unpublished data).

To test these hypotheses, lightly and darkly pigmented pools of F2 hybrid individuals from pairwise crosses of two lightly pigmented and two darkly pigmented lines of *D. americana* were genotyped at *tan*, *ebony*, and *yellow*. *tan* and *ebony* were chosen due to their involvement in the interspecies pigmentation divergence between *D. americana* and *D. novamexicana* as well as previous suggestion on their involvement in intraspecific pigmentation diversity within *D. americana* (Wittkopp et al., 2009). *yellow* was chosen due to its known role in *Drosophila* pigmentation evolution in multiple species groups (Gompel, Prud'homme, Wittkopp, Kassner, & Carroll, 2005; Jeong, Rokas, & Carroll, 2006; Ordway, Hancuch, Johnson, Wiliams, & Rebeiz, 2014; Prud'homme et al., 2006). Additionally, in this system, *yellow* serves as a control since prior study has not implicated *yellow* in pigmentation divergence between *D. americana* and *D. novamexicana* (Wittkopp et al., 2009).

Materials and Methods

Selecting D. americana *lines to analyze*

To test the hypothesis that genetic heterogeneity exists in phenotypically similar populations of *D. americana*, lines isolated from the same geographic location with similar phenotypes were needed. However, not all independent isofemale lines isolated from a particular geographic location may have a different genetic basis for their similar phenotype – many lines may share an underlying genotype. If lines with the same genotype underlying their similar pigmentation phenotype are crossed to create an F1 hybrid population, and those F1 individuals reproduce to create a F2 population, we would not expect to see pigmentation variation in this population. Without different genes causing similar phenotypes, recombination in the reproducing F1 populations cannot bring together new combinations of alleles leading to greater pigmentation variation than seen in the individual parental lines. However, if each line has a unique genetic basis for the similar pigmentation (i.e. genetic heterogeneity is present), recombination can bring these alleles into the same individual. In this scenario, we expect to see greater variation in the pigmentation phenotypes seen in the F2 populations, especially in the pigmentation extremes with some F2 individuals showing a noticeably lighter or darker pigmentation phenotype than their parents.

To select *D. americana* lines exhibiting genetic heterogeneity, four light lines and four dark lines were selected for a series of test crosses (Table 5.1). In both the light and dark classes, two of the lines where from one geographic location and the other two were from a different geographic location. Each pairwise cross within the light classes and dark classes were set, allowed to produce an F1 hybrid population, and subsequently crosses to produce an F2 population. The resulting 24 populations of F2 individuals (12 from darkly pigmented lines, 12

from lightly pigmented lines) were qualitatively divided into phenotypic classes that could be reliably separated from one another. Most crosses produced a similar number of phenotypic classes (4-5 different classes), suggesting that the lines selected have some amount of genetic heterogeneity. For the remainder of the experiment, the two lightly pigmented isofemale lines of *D. americana* isolated from the same geographic location (Nebraska) and two darkly pigmented isofemale lines of *D. americana* isolated from the same geographic location (Ohio) were used. Since gene flow between individuals in the same geographic location is higher than those isolated from different locations, these pairings and the suggested genetic heterogeneity are particularly interesting from an evolutionary biology perspective. The specific lines chosen are: WS07.12 (dark), WS07.16 (dark), NN07.10 (light), NN07.18 (light); throughout this chapter, these will be referred to as D1 (WS07.12), D2 (WS07.16), L1 (L1NN07.10), and L2 (NN07.18).

Table 5.1: *D. americana* strains evaluated for potential genetic heterogeneity. Geographical location from which the strain was isolated, the name of the strain, and pigmentation phenotype (light or dark) are listed.

Location	Name	Pigmentation
Duncan, NE	DN0748x37	Light
	DN0728x21	Light
Niobrara, NE	NN07.10	Light
	NN07.18	Light
Killbuck, OH	WS07.12	Dark
	WS07.16	Dark
Ottawa, OH	OR07.06	Dark
	OR01.92	Dark

Generating a population of F2 individuals with diverse pigmentation phenotypes

The two lightly pigmented isofemale lines of *D. americana* isolated from the same geographic location (Nebraska) and two darkly pigmented isofemale lines of *D. americana* isolated from the same geographic location (Ohio) were crossed in all pairwise and reciprocal combinations to produce 12 different F2 populations. These F2 individuals exhibit a pigmentation phenotypes,

including classes darker and lighter than either parental lines. F2 individuals from each cross were first sorted by sex, then each sex was divided into three pigmentation classes: light, medium, and dark. These classes were relative to each cross, thus a distinction of light from one cross may not have the same pigmentation intensity as the light class from a different cross. From each cross, the lightest 25 male individuals and 25 darkest male individuals were collected for subsequent analysis.

Obtaining genomic DNA from pools of pigmentation extremes

To obtain genomic DNA from the light and dark male pools for each cross, we utilized Qiagen's DNeasy Blood & Tissue kit. Specifically, we used the supplementary protocol for purification of genomic DNA from whole insects using a disposable microtube pestle (https://www.qiagen.com/us/resources/resourcedetail?id=cabd47a4-cb5a-4327-b10d-d90b8542421e&lang=en) This method of genomic DNA extraction was chosen for its ability to generate high quality DNA suitable for possible future genome-wide studies in this system. Briefly, 25 *D. americana* adult flies (approximately 50mg total weight) were homogenized together in PBS using a disposable microtube pestle. This homogenized sample was used as the starting material for the standard DNeasy Blood and Tissue kit protocol, which uses a buffer system to lyse cells and selectively bind DNA to a column membrane. A series of centrifugation washes on the column removes contaminants and purifies the eluted DNA.

Designing pyrosequencing assays for genotyping at tan, ebony, *and* yellow

As a first step to investigating the phenotypic divergence and genetic heterogeneity in the *D. americana* lines chosen for study, a candidate gene approach was employed for genotyping.

Specifically, the pigmentation genes *tan*, *ebony*, and *yellow* were chosen for genotyping the F2 pigmentation extreme pools. These genes were chosen because of their known involvement in the biosynthetic pathway of pigmentation in *Drosophila* as well as their role in pigmentation divergence within and between *Drosophila* species (reviewed in Wittkopp & Massey, 2017). In order to genotype the different lines at *tan, ebony*, and *yellow*, differences in sequence between the four *D. americana* lines needed to be discovered. To this end, noncoding sequences within or adjacent to *tan, ebony,* and *yellow* were Sanger sequenced in each line. Noncoding sequences were chosen since they are more likely to harbor sequence differences than protein-coding sequences, especially within a single species. SNPs were identified that could be used to identify each the allele from each line in *tan, ebony,* and *yellow.*

Pyrosequencing assays to quantify the relative frequency of each allele in the F2 lightly and darkly pigmented pools were designed for each gene. For *ebony*, a single pyrosequencing reaction can be used to for all crosses; for *tan* and *yellow*, more than one assay was required due to SNPs not being in close enough proximity to one another. In each case, PCR primers were used to amplify the sequences analyzed (Table 5.2). With numerous pyrosequencing assays required, efforts were made to minimize the economic impact of ordering biotinylated primers for each assay; instead, a universal pyrosequencing primer strategy was employed (Aydin, Toliat, Bahring, Becker, & Nurnberg, 2006; Guo & Milewicz, 2007; Pacey-Miller & Henry, 2003; Royo, Hidalgo, & Ruiz, 2007). For each assay, one of the primers was ordered with a tail (underlined sequence in Table 5.2) complementary to either the forward or reverse biotinylated universal pyrosequencing primer.

Table 5.2: PCR amplification primers and pyrosequencing primers used to measure allele frequency in F2 hybrid extreme pigmentation pools.

gene	Forward primer	Reverse primer	Pyrosequencing primer
tan	CTACTTTTTTGCCATTTCGTGACC	TAGCAGGATACGACTATCTTGTTTTTCCGGCTCAAAGCGA	CTTCACTCAAAACTAACACT
tan	CGCTTTGAGCCGGAAAAAC	TAGCAGGATACGACTATCGAGCAGGAGTGGGTCCAGA	TGGCAGGTGAGCGGG
tan	GTGACGTACTAGCAACGCCCAGTTCCTGGCTTAACCTT	GCCCTTGATATTGGCCACTT	CAAGTGAGTGATTGTTATAT
ebony	GTGACGTACTAGCAACGTTAAAGGTAGTTCCATTAGACTTTG	CGAACTTCCAACTTCTAGAGC	TATATGAATCGAACGAT
yellow	GTGACGTACTAGCAACGTTCAGTAAAAGGTTCTAGATCCAA	CTCTCTTTTCAATAATGCCTTTAATT	GTTGAGTCTATCGATAAC CGAAATCGATAAATATCGC

Universal forward primer: [Btn]5'-GTGACGTACTAGCAACG-3'
Universal reverse primer: [Btn]5'-TAGCAGGATACGACTATC-3'

Genotyping at tan, ebony, *and* yellow *using pyrosequencing*

For each pyrosequencing assay, PCR was performed using both the forward and reverse primers listed in Table 5.2 along with the appropriate universal primer. In addition to the 24 samples of genomic DNA extracted from the dark and light pools of each F2 cross, genomic DNA extracted from each parental line (D1, D2, L1, L2) and a pool of 25 F1 hybrids from each pairwise combination. These samples serve as controls – the parental lines to ensure the assays accurately identify the SNPs used and the F1 lines to control for any PCR bias as these pools should have a 1:1 ratio of each line's allele. In one instance (L1, D2), a mixture of parental lines needed to be used instead of F1 hybrid individuals due to a failure to reproduce; while this is not ideal due to potential differences in body size between the individuals selected, care was taken to select 13 individuals from each line with similar size and were mixed prior to genomic DNA extraction. For each F2 population, a ratio was calculated comparing the two alleles present, such that a value equal to one indicates equal representation of the two alleles in the population. Each of

these alleles was corrected for possible PCR bias using the F1 hybrid DNA control where the ratio of alleles is 1:1.

Data analysis

Allele frequency for the light and dark pigmentation pools for each cross and technical replicate was calculated and subsequently analyze in R v3.4.0 (RCoreTeam, 2016). The log2 ratio of allele frequency was taken to center the data around zero. The log2 ratio of allele frequency was fitted to a linear model to test the effect of pigmentation classification (light, dark). A t-test was performed to identify significant differences in the allele frequencies between light and dark pools. A Bonferroni correction was used to obtain a conservative significance cutoff of $p<0.001471$.

Results

For each cross, the F2 population of flies were separated into the lightest, intermediate, and darkest pigmentation classes. The 25 lightest and 25 darkest flies were used for genotyping at *tan, ebony,* and *yellow.* In these phenotypic extreme pools of F2 hybrid individuals, if a gene contributes to the pigmentation variation between two lines of *D. americana*, the expectation is that the allele frequency will be different between the lightly pigmented and darkly pigmented pigmentation pools. However, if a gene does not contribute to pigmentation variation, the expectation is that the two alleles will be at approximately equal levels in both the lightly pigmented and darkly pigmented pools. Using these expectations, the contribution of *tan*, *ebony*, and *yellow* in intraspecies pigmentation in *D. americana* was evaluated. The results from these experiments are summarized in Table 5.3 and discussed below.

Table 5.3: Associations between relative allele frequency and pigmentation in F2 hybrid pigmentation extreme pools. p-values for each comparison with significant values shown in red text (Bonferroni corrected significance cutoff of $p < 0.001471$, p-values obtained from t-test).

		Is pigmentation associated with:		
Allele 1	**Allele 2**	***tan*?**	***ebony*?**	***yellow*?**
D1	D2	4.79E-04	2.04E-03	0.617
D2	D1	0.924	3.55E-03	0.179
D1	L1	0.024	0.017	4.09E-05
L1	D1	0.036	0.138	0.291
D1	L2	0.044	1.07E-04	0.088
L2	D1	0.012	5.03E-03	0.143
D2	L1	5.08E-03	3.26E-03	0.036
L1	D2	8.70E-03	0.017	0.258
D2	L2	0.05	3.38E-03	0.394
L2	D2	0.225	0.095	0.589
L1	L2	1.50E-10	5.89E-03	-
L2	L1	0.033	0.095	-

For the comparison between the two dark lines of *D. americana* (D1, D2), representative flies from both parental lines, as well as the lightest and darkest pigmentation pools from the reciprocal crosses are shown in Figure 5.2. When considering this comparison, only *ebony* was significantly associated with pigmentation, and only in one of the reciprocal crosses (Table 5.3, Figure 5.2, Figure 5.3). In the significant cross, the D2 allele was enriched in both the light and dark pigmentation pools, but significantly more so in the dark pigmentation extreme pool of individuals. *yellow* and *tan* were not significantly associated with pigmentation differences. Differences in allele frequency at *ebony* were close to the statistical significance cutoff and may show a significant association with increased power of detection via additional replication.

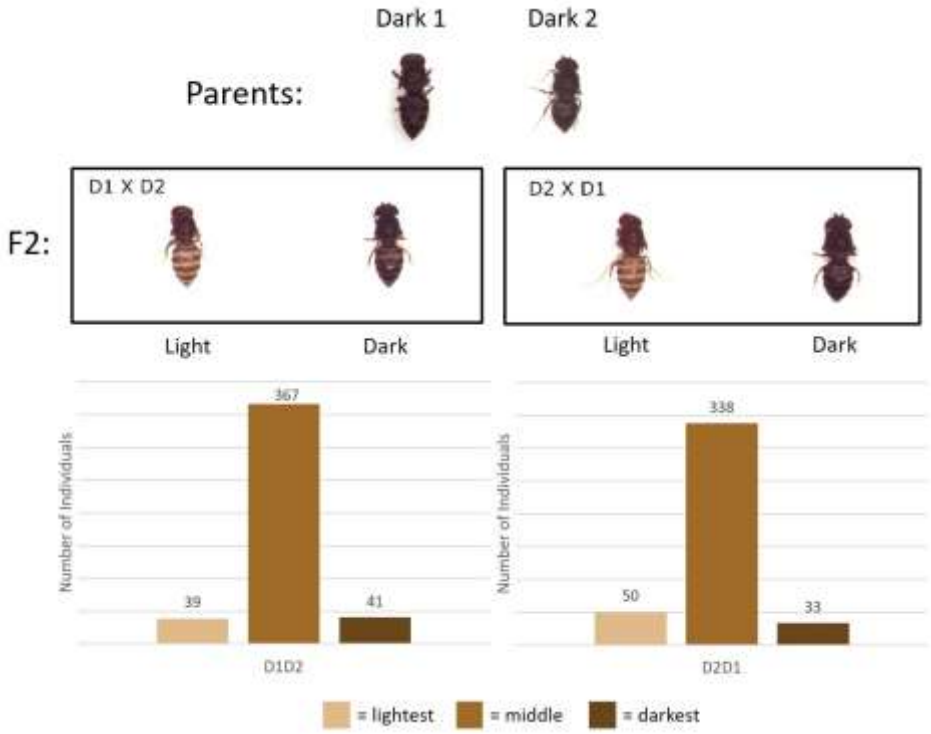

Figure 5.2: Representative images of D1 and D2 parental lines and F2 hybrid populations. Representative females are shown for the parental lines (Dark 1, Dark 2) as well as both reciprocal crosses (D1 x D2, D2 x D1); for each F2 cross, a representative from the lightest pigmentation class and darkest pigmentation class is shown. Histograms below the F2 representative individuals show the frequency of classification into lightest, middle, and darkest pigmentation classes for the F2 populations.

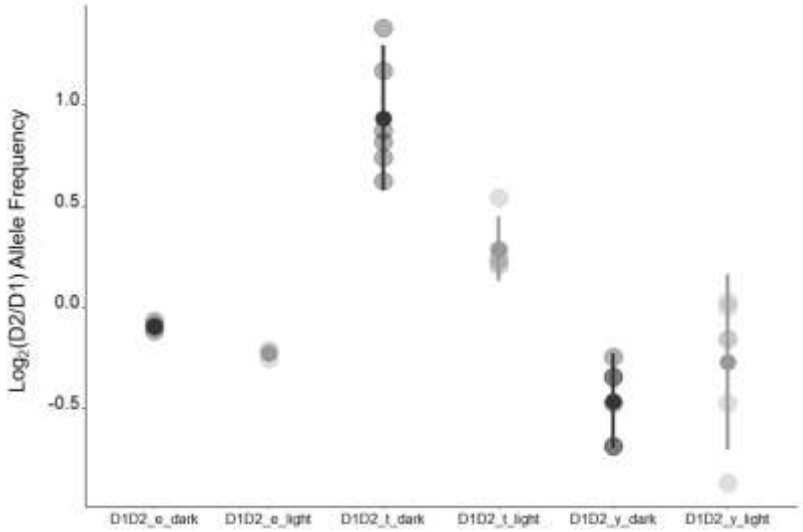

Figure 5.3: Allelic frequency in D1 x D2 pigmentation extremes. Log_2 ratio of allelic abundance between D1 and D2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D2 allele, while values below zero indicate enrichment of the D1 allele.

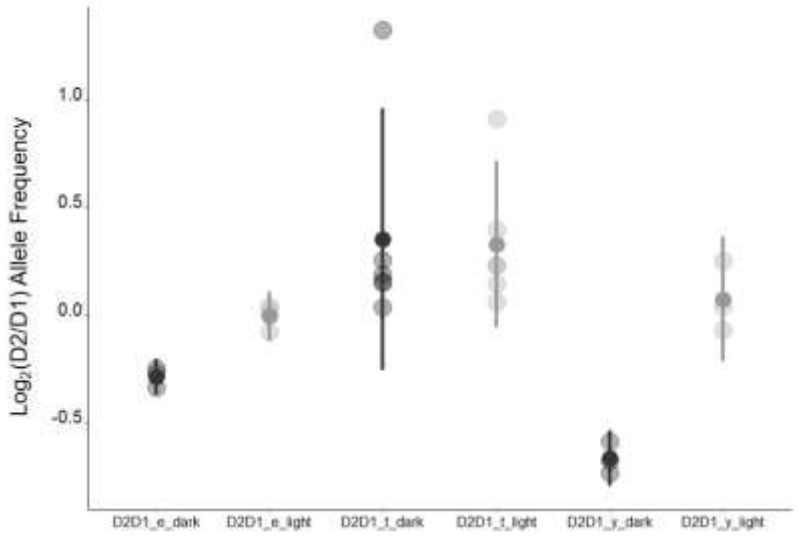

Figure 5.4: Allelic frequency in D2 x D1 pigmentation extremes. Log_2 ratio of allelic abundance between D1 and D2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D2 allele, while values below zero indicate enrichment of the D1 allele.

When evaluating the Dark 1 (D1) line of *D. americana* and Light 1 (L1) line of *D. americana* (Figure 5.5), only *yellow* was significantly associated with pigmentation, and only in one of the reciprocal crosses (Table 5.3, Figure 5.6, Figure 5.7). In the significant cross, the D1 allele of *yellow* is enriched in the light pigmentation pool, while the L1 allele of *yellow* was enriched in the dark pigmentation pool. This is counterintuitive considering the pigmentation phenotypes of the parental lines and the role of *yellow* in making dark pigments. *ebony* and *tan* were not significantly associated with pigmentation differences.

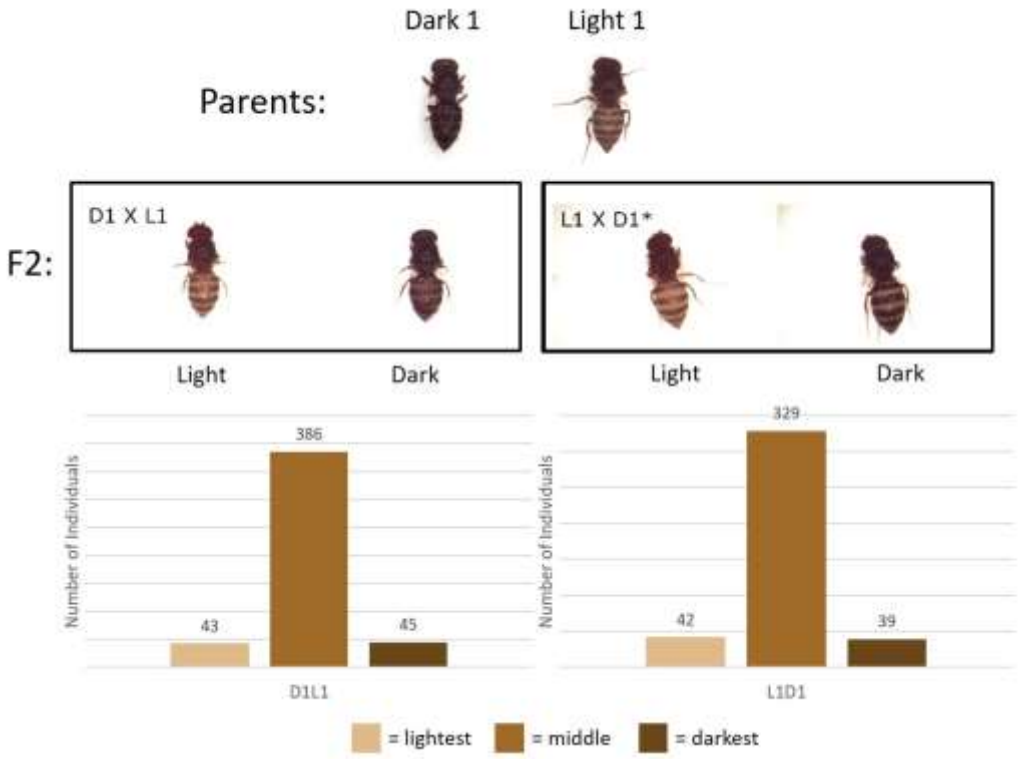

Figure 5.5: Representative images of D1 and L1 parental lines and F2 hybrid populations. Representative females are shown for the parental lines (Dark 1, Light 1) as well as both reciprocal crosses (D1 x L1, L1 x D1); for each F2 cross, a representative from the lightest pigmentation class and darkest pigmentation class is shown. Histograms below the F2 representative individuals show the frequency of classification into lightest, middle, and darkest pigmentation classes for the F2 populations.

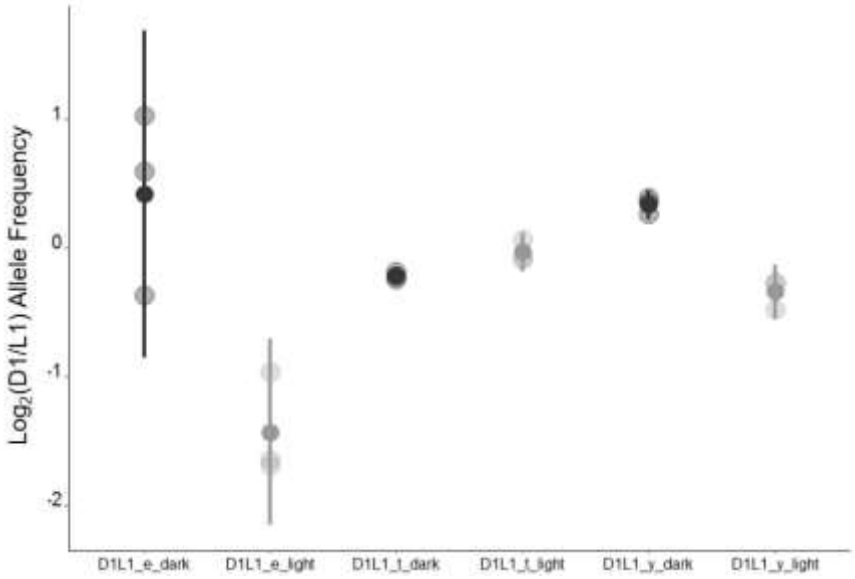

Figure 5.6: Allelic frequency in D1 x L1 pigmentation extremes. Log_2 ratio of allelic abundance between D1 and L1 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D1 allele, while values below zero indicate enrichment of the L1 allele.

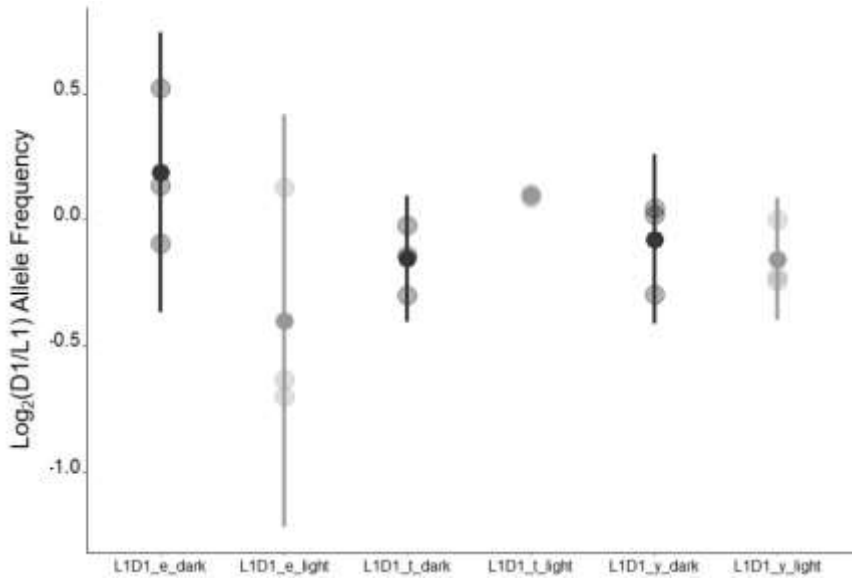

Figure 5.7: Allelic frequency in L1 x D1 pigmentation extremes. Log_2 ratio of allelic abundance between D1 and L1 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D1 allele, while values below zero indicate enrichment of the L1 allele.

In the comparison between Dark 1 (D1) and Light 2 (L2), only *ebony* was significantly associated with pigmentation (Table 5.3), and only in one of the reciprocal crosses (Figure 5.8, Figure 5.9, Figure 5.10). In this instance, the L2 allele of *ebony* was significantly enriched in the light pigmentation pools, while the D1 allele of *ebony* was significantly enriched in the dark pigmentation extreme pool of individuals. Since *ebony* catalyzes a reaction leading to light pigment formation, this follows expectations based on gene function and the parental phenotypes. For the reciprocal cross, the difference in allele frequency at *ebony* was close to the statistical significance cutoff and may show a significant association with increased power of detection via additional replication. *yellow* and *tan* were not significantly associated with pigmentation differences.

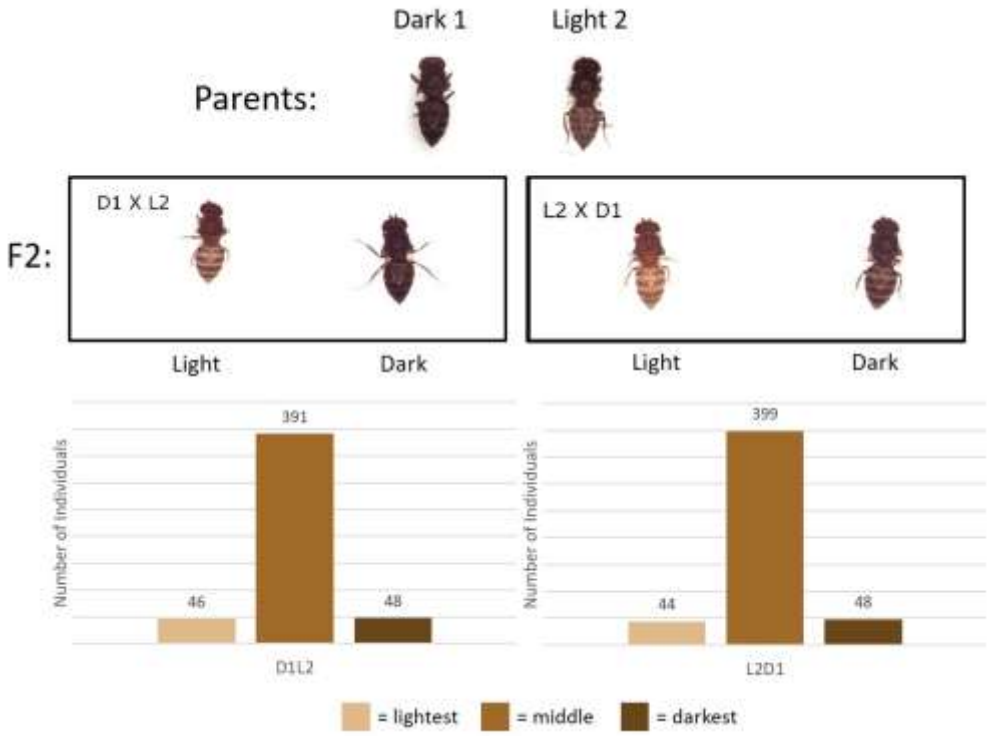

Figure 5.8: Representative images of D1 and L2 parental lines and F2 hybrid populations. Representative females are shown for the parental lines (Dark 1, Light 1) as well as both reciprocal crosses (D1 x L2, L2 x D1); for each F2 cross, a representative from the lightest pigmentation class and darkest pigmentation class is shown. Histograms below the F2 representative individuals show the frequency of classification into lightest, middle, and darkest pigmentation classes for the F2 populations.

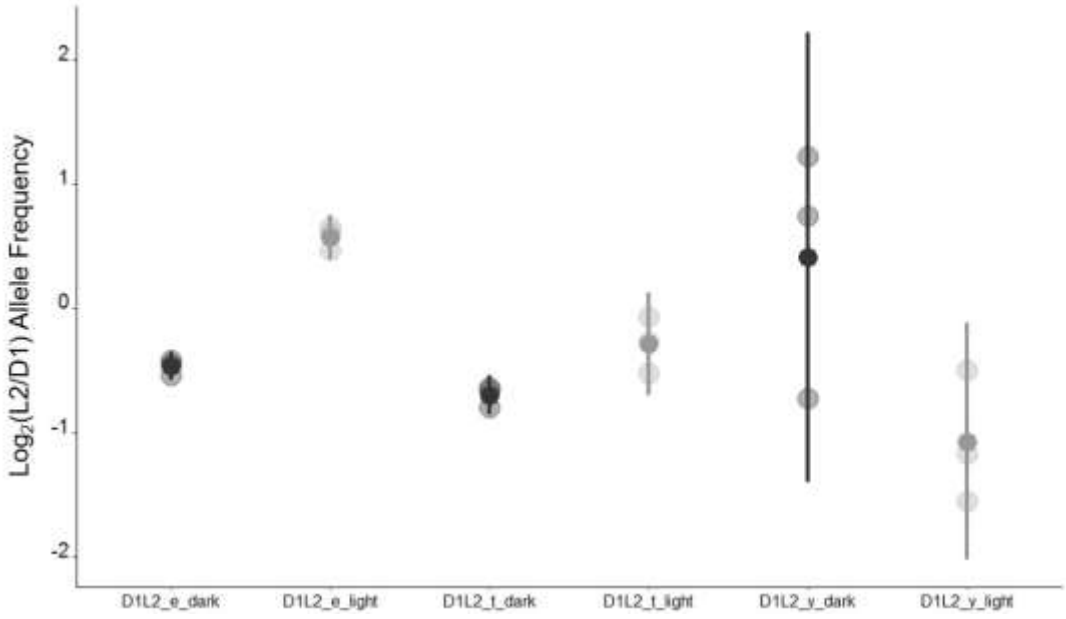

Figure 5.9: Allelic frequency in D1 x L2 pigmentation extremes. Log_2 ratio of allelic abundance between D1 and L2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D1 allele, while values below zero indicate enrichment of the L2 allele.

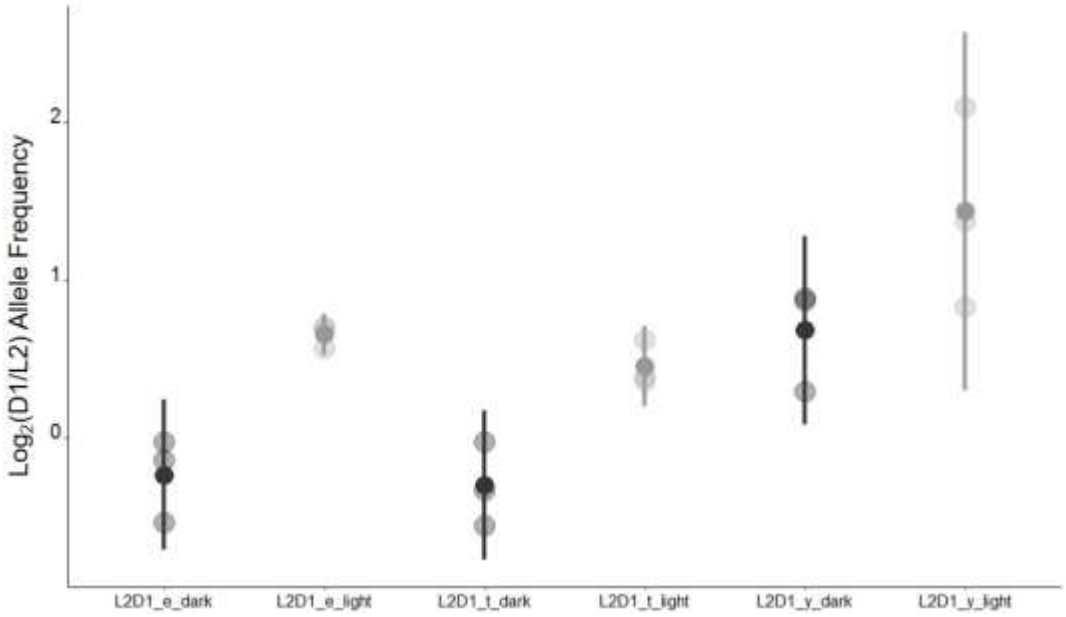

Figure 5.10: Allelic frequency in L2 x D1 pigmentation extremes. Log_2 ratio of allelic abundance between D1 and L2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D1 allele, while values below zero indicate enrichment of the L2 allele.

In the comparison between two light lines of *D. americana* (L1, L2), only *tan* was significantly associated with pigmentation (Table 5.3), and only in one of the reciprocal crosses (Figure 5.11, Figure 5.12, Figure 5.13). In this instance, the L1 allele of *tan* was significantly enriched in the light pigmentation pools, while the L2 allele of *tan* was significantly enriched in the dark pigmentation extreme pool of individuals. *yellow* and *tan* were not significantly associated with pigmentation differences.

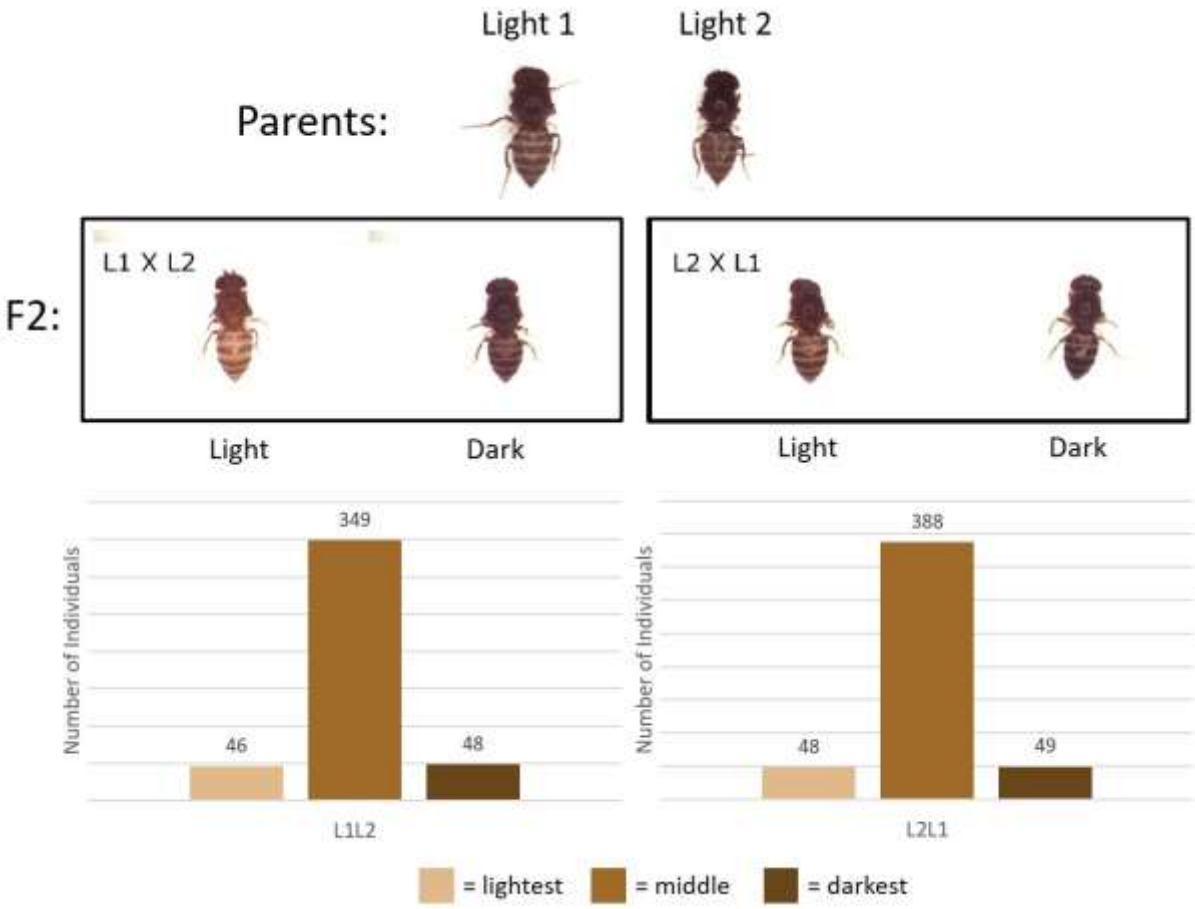

Figure 5.11: Representative images of L1 and L2 parental lines and F2 hybrid populations. Representative females are shown for the parental lines (Dark 1, Light 1) as well as both reciprocal crosses (L1 x L2, L2 x L1); for each F2 cross, a representative from the lightest pigmentation class and darkest pigmentation class is shown. Histograms below the F2 representative individuals show the frequency of classification into lightest, middle, and darkest pigmentation classes for the F2 populations.

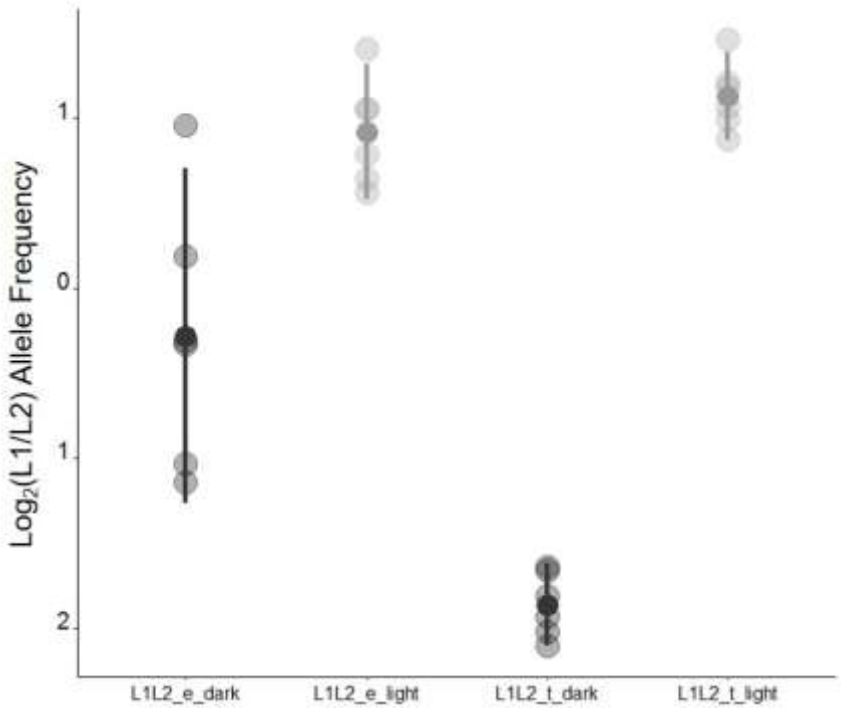

Figure 5.12: Allelic frequency in L1 x L2 pigmentation extremes. Log_2 ratio of allelic abundance between L1 and L2 alleles at *ebony* (e) and *tan* (t). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the L1 allele, while values below zero indicate enrichment of the L2 allele.

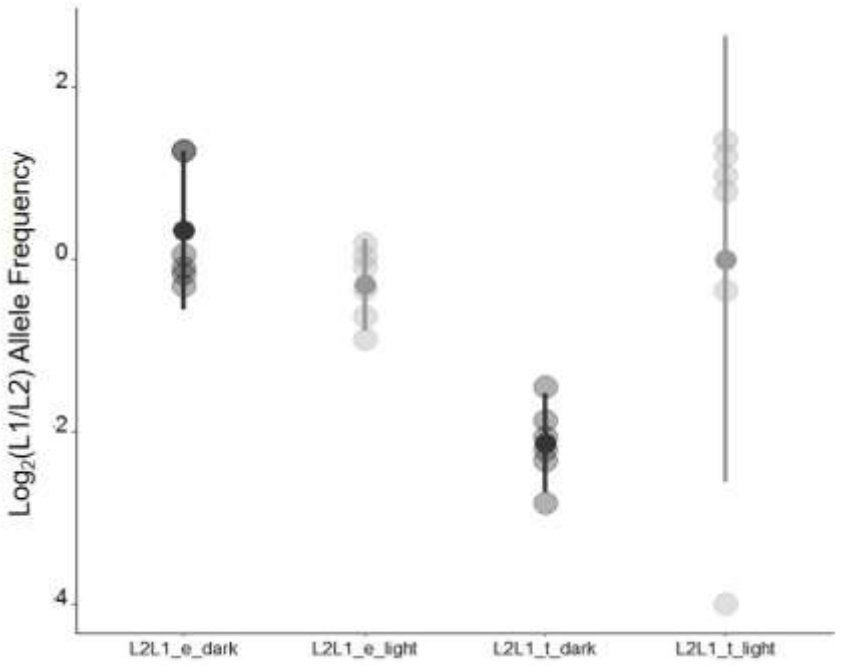

Figure 5.13: Allelic frequency in L2 x L1 pigmentation extremes. Log_2 ratio of allelic abundance between L1 and L2 alleles at *ebony* (e) and *tan* (t). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from

light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the L1 allele, while values below zero indicate enrichment of the L2 allele.

For the remaining two comparisons (D2 & L1, D2 & L2) there were no significant association found at *tan*, *ebony*, or *yellow*. The results for D2 and L1 are shown in Figures 5.14-5.16. Results for D2 and L2 are shown in Figures 5.17-5.19.

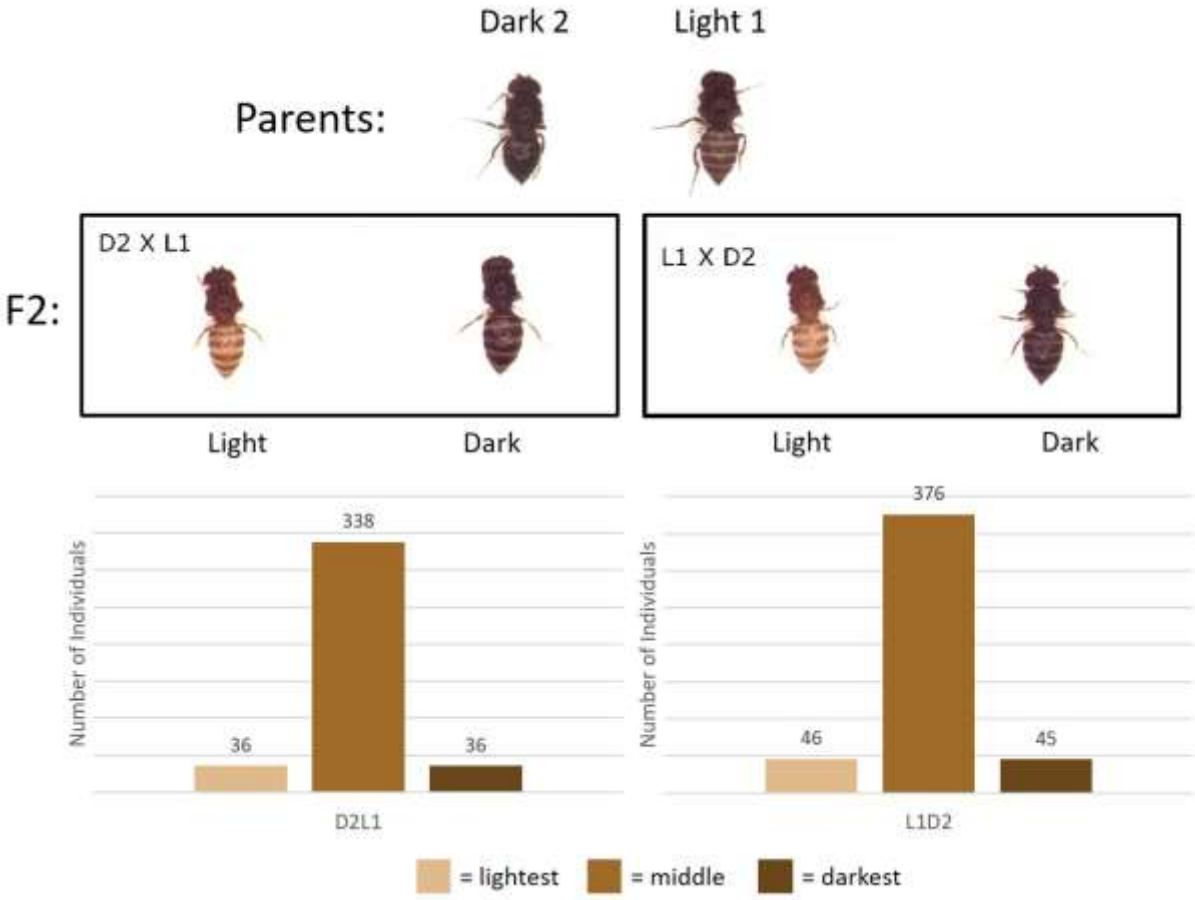

Figure 5.14: Representative images of D2 and L1 parental lines and F2 hybrid populations. Representative females are shown for the parental lines (Dark 1, Light 1) as well as both reciprocal crosses (D2 x L1, L1 x D2); for each F2 cross, a representative from the lightest pigmentation class and darkest pigmentation class is shown. Histograms below the F2 representative individuals show the frequency of classification into lightest, middle, and darkest pigmentation classes for the F2 populations.

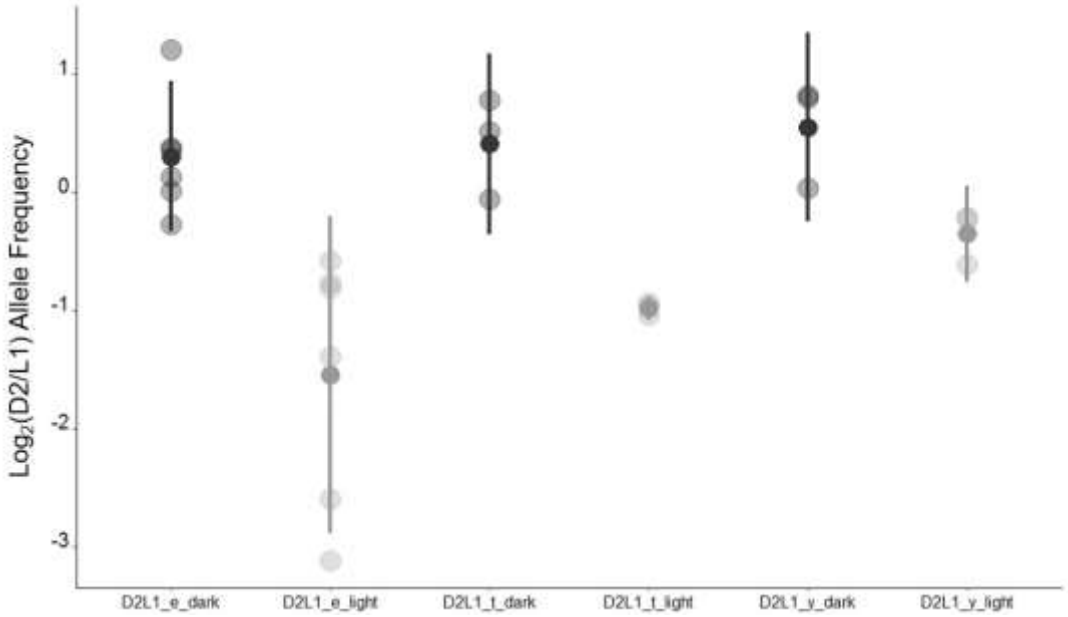

Figure 5.15: Allelic frequency in D2 x L1 pigmentation extremes. Log_2 ratio of allelic abundance between L1 and D2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the L1 allele, while values below zero indicate enrichment of the D2 allele.

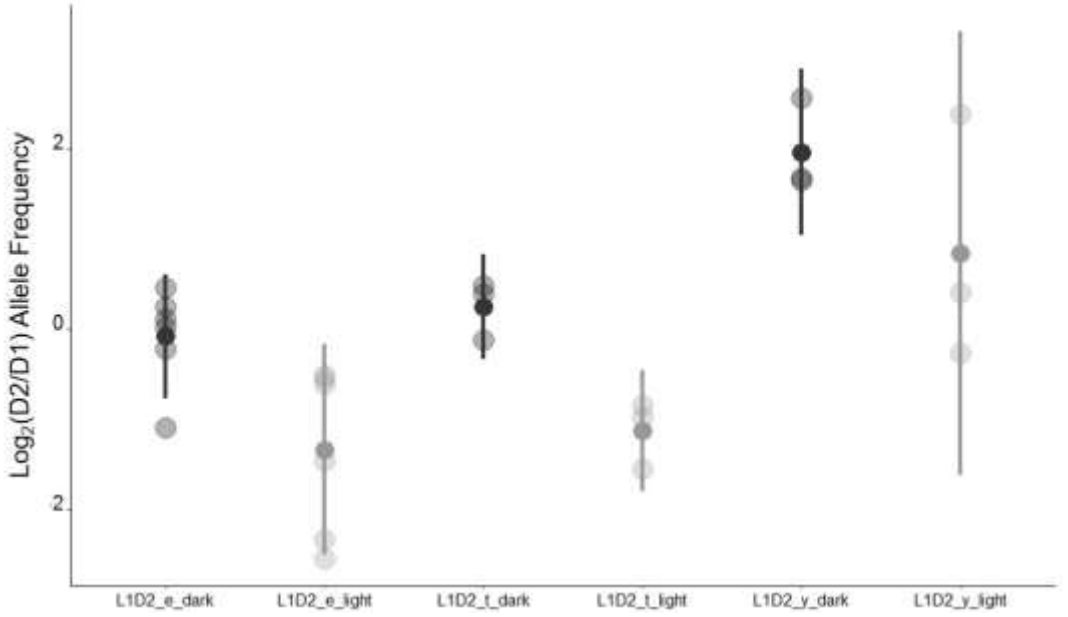

Figure 5.16: Allelic frequency in L1 x D2 pigmentation extremes. Log_2 ratio of allelic abundance between L1 and D2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the L1 allele, while values below zero indicate enrichment of the D2 allele.

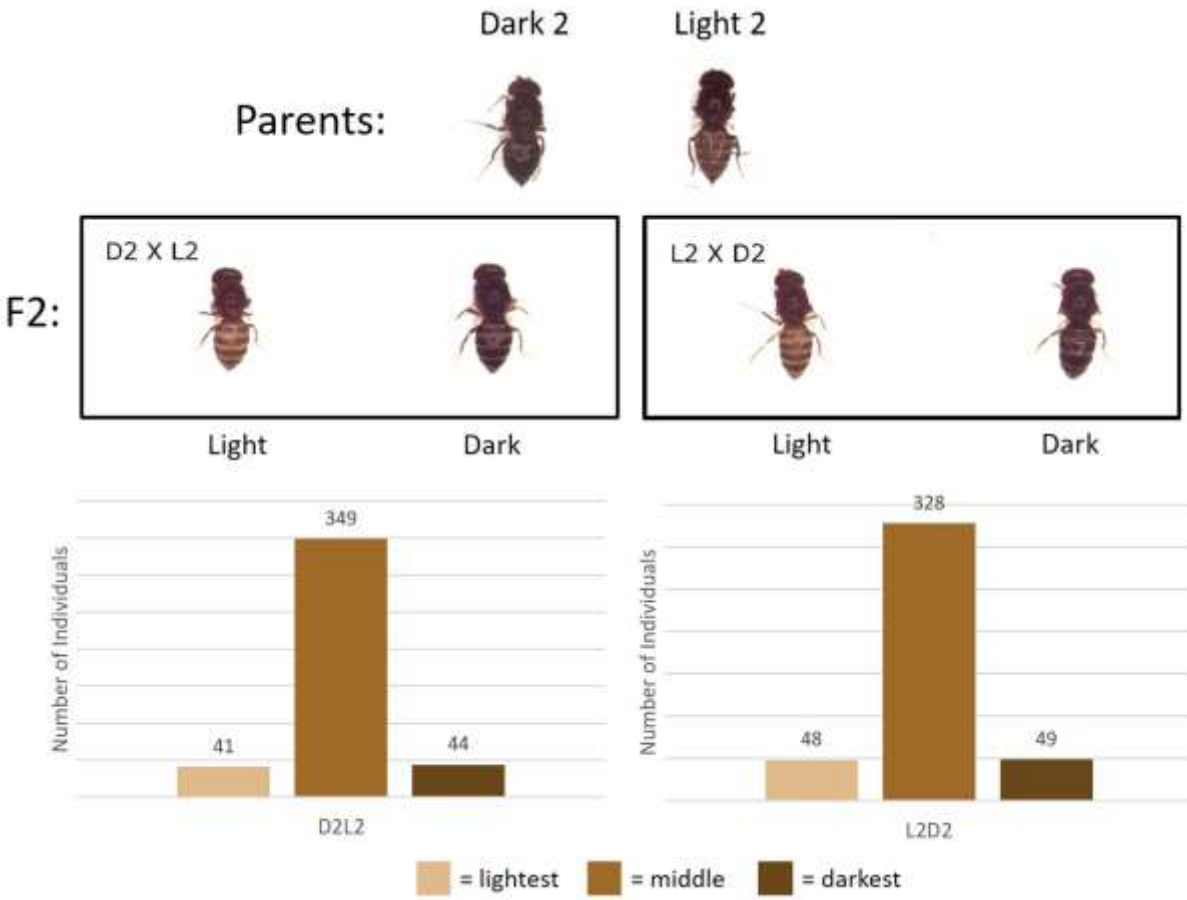

Figure 5.17: Representative images of D2 and L2 parental lines and F2 hybrid populations. Representative females are shown for the parental lines (Dark 1, Light 1) as well as both reciprocal crosses (D2 x L2, L2 x D2); for each F2 cross, a representative from the lightest pigmentation class and darkest pigmentation class is shown. Histograms below the F2 representative individuals show the frequency of classification into lightest, middle, and darkest pigmentation classes for the F2 populations.

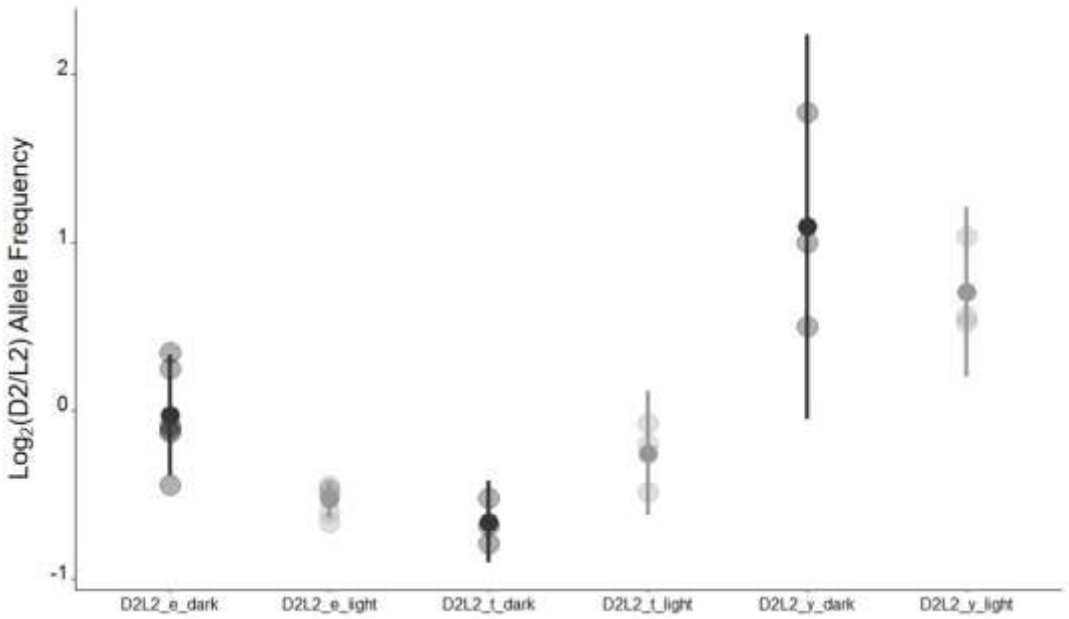

Figure 5.18: Allelic frequency in D2 x L2 pigmentation extremes. Log_2 ratio of allelic abundance between D2 and L2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D2 allele, while values below zero indicate enrichment of the L2 allele.

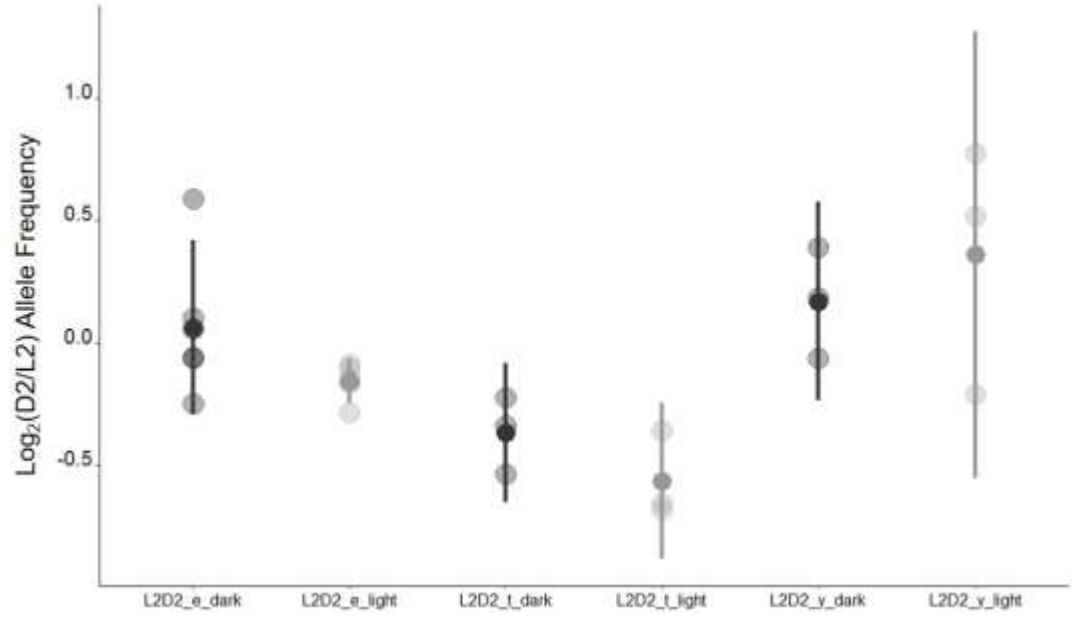

Figure 5.19: Allelic frequency in L2 x D2 pigmentation extremes. Log_2 ratio of allelic abundance between D2 and L2 alleles at *ebony* (e), *tan* (t), and *yellow* (y). Values from dark pools are shown in dark grey with mean allelic abundance shown in black. Values from light pigmentation pools are shown in lighter grey with mean allelic abundance shown in medium grey. Confidence intervals represent the Bonferroni significance cutoff value of $\alpha = 0.001417$ (99.999% confidence interval). Values above zero indicate an enrichment of the D2 allele, while values below zero indicate enrichment of the L2 allele.

Discussion

Three hypotheses were tested using the data set presented above: 1) *tan* and *ebony* contribute to intraspecies pigmentation variation in *D. americana*, 2) genetic variation at other genes also contributes to pigmentation variation in *D. americana*, and 3) genetic heterogeneity exists in *D. americana* such that phenotypically similar independent lines of *D. americana* have different underlying genotypes.

The contribution of tan *and* ebony *in* D. americana *pigmentation variation*

Both *tan* and *ebony* were associated with differences in pigmentation between F2 hybrid populations that were separated into light and dark pigmentation pools.

tan was associated with pigmentation differences between the two lightly pigmented lines of *D. americana* and between the two darkly pigmented lines of *D. americana.* In the case of the lightly pigmented lines, the L1 allele was present at a higher frequency in the lightly pigmented pool of F2 hybrid individuals and the L2 allele present at a higher frequency in the darkly pigmented pool. This suggests that the L1 allele of *tan* contributes to the light pigmentation phenotype exhibited by this line of *D. americana*, but that other genetic loci are responsible for the light pigmentation in L2. For the darkly pigmented lines, the D2 allele was enriched in both the light and dark pools, however, significantly more so in the dark pool. This result is surprising since it is expected that an allele contributing to pigmentation differences would be enriched in one phenotypic class and depleted in the other.

ebony was associated with differences in pigmentation between a light and dark line of *D. americana*, specifically D1 and L2. The L2 allele was enriched in the lightly pigmented pool and

the D1 allele was enriched in the darkly pigmented pool. This suggests that *ebony* contributes to the pigmentation divergence seen between these strains of *D. americana* in a manner consistent with *ebony* contributing to light pigmentation formation in pigmentation development in *Drosophila* (Wittkopp, True, & Carroll, 2002).

Overall, these results suggest that some of the same genes contribute to both inter- and intra-species pigmentation variation since *tan* and *ebony* are responsible for the majority of pigmentation divergence between *D. americana* and *D. novamexicana*. However, the lack of association with *tan* and/or *ebony* for many of the F2 hybrid populations studied suggests that other genes have a significant contribution to pigmentation variation within *D. americana* but not between *D. americana* and *D. novamexicana.*

Involvement of other genes in pigmentation diversity within D. americana

As a pilot experiment to investigate if genes other than *tan* and *ebony* were involved in the pigmentation divergence seen within *D. americana*, the pigmentation gene *yellow* was genotyped in the F2 hybrid populations. *yellow* is known to be involved with pigmentation evolution in multiple *Drosophila* species (Gompel et al., 2005; Jeong et al., 2006; Ordway et al., 2014; Prud'homme et al., 2006), but not between *D. americana* and *D. novamexicana* (Wittkopp et al., 2009). Thus, *yellow* represents a reasonable candidate for contribution to intraspecies variation, yet one that is not expected to contribute to variation in this species. Surprisingly, *yellow* was found to be associated with pigmentation in an F2 population from D1 and L1. The D1 allele was enriched in the light pigmentation pool, which is also surprising given the role of *yellow* in dark pigmentation development and the phenotypes of the lines examined. This result

gives an unclear conclusion about how or if *yellow* contributes to pigmentation variation within *D. americana.*

Pigmentation was not associated with *tan, ebony,* or *yellow* in two comparisons in this experiment (D2 vs L1, D2 vs L2) despite obvious differences in pigmentation between the F2 hybrid pigmentation pools. This suggests that the genetic variation underlying these phenotypes is not linked to *tan*, *ebony*, or *yellow*. Future study of these F2 populations using a genome-wide approach such as RAD-Seq would help identify additional loci involved in pigmentation diversity within *D. americana*. In this type of study, similar logic would be utilized, with the expectation that allelic frequency link to loci contributing to pigmentation divergence would be significantly different between light and dark pools of F2 hybrid individuals.

Genetic heterogeneity in D. americana

The F2 populations created between *D. americana* lines exhibiting similar pigmentation phenotypes represent a unique opportunity to explore genetic heterogeneity in these populations. If the same genotypes underlie the similar pigmentation phenotypes, I would not expect to see variation in pigmentation in the F2 population equal to or greater than the variation seen in the individual parental lines. In addition, I expect the ratio of alleles in the F2 populations to be near 1:1 and not show significant differences in allele frequency between the light and dark pigmentation pools. However, if genetic heterogeneity is present in this population, I expect to see a wider pigmentation phenotypes in the F2 population, as well as a significant difference in allele frequency in the F2 pigmentation pools at loci that underlie the pigmentation phenotype in one strain of *D. americana* but not the other.

In both comparisons between the lines of *D. americana* with similar phenotypes, *tan* showed an association with pigmentation differences. This suggests that *tan* contributes to the pigmentation phenotypes of these lines despite their similarity in appearance. This result supports the hypothesis that genetic heterogeneity exists in *D. americana*.

Acknowledgments

This work would not have been possible without: Lisa Sramkoski and Wesley McLaughlin for their contributions to the foundation of this work; Ali Farhat for assistance with generation and maintenance of the *Drosophila americana* populations used in this work; Jennifer Lachowiec and Fabien Duveau for contribution to pyrosequencing protocols, assay design, and statistical analysis.

References

Aydin, A., Toliat, M. R., Bahring, S., Becker, C., & Nurnberg, P. (2006). New universal primers facilitate Pyrosequencing. *Electrophoresis, 27*(2), 394-397. doi:10.1002/elps.200500467

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., & Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature, 433*(7025), 481-487. doi:10.1038/nature03235

Guo, D. C., & Milewicz, D. M. (2007). Universal primer applications for pyrosequencing. *Methods Mol Biol, 373*, 57-62. doi:10.1385/1-59745-377-3:57

Jeong, S., Rokas, A., & Carroll, S. B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution. *Cell, 125*(7), 1387-1399. doi:10.1016/j.cell.2006.04.043

Massey, J. H., & Wittkopp, P. J. (2016). The Genetic Basis of Pigmentation Differences Within and Between Drosophila Species. *Curr Top Dev Biol, 119*, 27-61. doi:10.1016/bs.ctdb.2016.03.004

Ordway, A. J., Hancuch, K. N., Johnson, W., Wiliams, T. M., & 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*, *392*(2), 431-440. doi:10.1016/j.ydbio.2014.05.023

Pacey-Miller, T., & Henry, R. (2003). Single-nucleotide polymorphism detection in plants using a single-stranded pyrosequencing protocol with a universal biotinylated primer. *Anal Biochem*, *317*(2), 166-170.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., . . . Carroll, S. B. (2006). Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*, *440*(7087), 1050-1053. doi:10.1038/nature04597

Royo, J. L., Hidalgo, M., & Ruiz, A. (2007). Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. *Nat Protoc*, *2*(7), 1734-1739. doi:10.1038/nprot.2007.244

Wittkopp, P. J., Smith-Winberry, G., Arnold, L. L., Thompson, E. M., Cooley, A. M., Yuan, D. C., . . . McAllister, B. F. (2011). Local adaptation for body color in Drosophila americana. *Heredity (Edinb)*, *106*(4), 592-602. doi:10.1038/hdy.2010.90

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., . . . Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science*, *326*(5952), 540-544. doi:10.1126/science.1176980

Wittkopp, P. J., True, J. R., & Carroll, S. B. (2002). Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns. *Development*, *129*(8), 1849-1858.

Chapter 6

Conclusions

The incredible amount of phenotypic diversity seen throughout the tree of life has arisen from a common ancestor over millions of years. Over time, heritable changes in DNA combined with adaptive (i.e. natural selection) and non-adaptive processes (i.e. genetic drift) have contributed to the production of novel phenotypes. However, not all genetic mutations impact phenotypes, nor is it straightforward to understand which mutations have an effect, what effect they will have on phenotypic output, or what molecular mechanism they use to influence phenotypes. Understanding this genetic basis of phenotypic diversity has been a fundamental goal of biology.

A common distinction between mutations influencing phenotypic diversity is coding or noncoding; that is, does the mutation occur in a DNA sequences that encodes the amino acid sequence for a protein (coding) or in a DNA sequence that does not (noncoding). When considering these two classifications of mutation, our ability to predict the effect of a coding change is much greater given our understanding of the genetic code. By looking at the sequence of coding DNA, one can infer the specific effect a given mutation (synonymous substitution, nonsynonymous substitution, frameshift, nonsense) will have on the resulting amino acid sequence of the protein. Our ability to evaluate the potential impact of mutations in noncoding DNA sequences is much less powerful.

While noncoding DNA sequences can have many roles in an organism, one of the most well studied is controlling transcriptional level gene expression. *cis*-regulatory elements, such as promoters and enhancers, are noncoding DNA sequences that control when, where, and to what extent a gene is turned on/off. Any given gene can have multiple *cis*-regulatory elements that work in concert to produce the expression pattern of that gene. Often, these independent *cis*-regulatory elements control a particular aspect of gene expression, such as driving gene expression in a particular developmental stage or location. This modularity is one factor contributing to the *cis*-regulatory hypothesis (Stern & Orgogozo, 2008) which states that mutations causing phenotypic diversity in morphological traits are most likely to arise in *cis*-regulatory sequences. Empirical data supports this hypothesis, especially when considering phenotypic change between different species (Stern & Orgogozo, 2008). However, questions remain about the how changes in *cis*-regulatory sequences influence the production of divergent phenotypes, both within and between species. In the work presented here, pigmentation divergence between *Drosophila americana* and *Drosophila novamexicana*, as well as within *D. americana*, was used to study this and other questions related to the role of noncoding sequences in phenotypic evolution. This system also provides a unique opportunity to test similarities and differences in phenotypic evolution between and within species.

Sensitivity of allelic divergence to genomic position

To identify genetic variants underlying phenotypic difference, as well as test alleles for divergent function, researchers often utilize transgenic animals to compare the function of alleles. In many model organisms, including *Drosophila*, targeted integration makes it possible to evaluate the effect of multiple alleles at the same genomic location. When using this approach,

it is generally assumed that position effects (Sturtevant, 1925; Wilson, Bellen, & Gehring, 1990) are controlled for since each allele evaluated in the same genomic context. Using *tan* alleles from *D. americana* and *D. novamexicana* inserted at five different genomic locations in *D. melanogaster*, this assumption was formally tested. I found that the relative effects of these *tan* alleles varied among the genomic locations tested. Three of the five genomic locations allowed functional differences to be detected, while the remaining two locations did not. One of those two simply silenced both transgenes; the other, however, allowed sufficient expression to rescue a mutant phenotype yet failed to detect the functional differences between *D. americana* and *D. novamexicana tan* alleles seen at the remaining genomic locations. This finding has implications for the field since it suggests that functional divergence between alleles could be undetected if they are examined in only one genomic location.

While this work is not revolutionary in terms of understanding how changes in phenotypes arise over time, it is an important contribution to molecular biology study. With transgenic studies being a gold standard in testing for functional differences between alleles, it is important that researchers understand the potential impact their experimental design has on their conclusions. Using a single genomic location to control for position effects is a common practice, yet is not valid in all cases. From the work presented here, there appears to be an expression threshold needed in order to detect functional differences between different alleles. Future studies could examine these and additional landing sites to see if these conclusions hold true for additional transgenes and additional genomic locations.

Evaluating the contribution of tan *to pigmentation divergence between* D. americana *and* D. novamexicana *using transgenic alleles*

Transgenic analysis was used to evaluate the effect of *D. americana* and *D. novamexicana tan* alleles on pigmentation in a *D. melanogaster* host. I found that pigmentation differences driven by the *D. americana* and *D. novamexicana tan* alleles in *D. melanogaster* were robust to minor variations in experimental conditions and conformed to expectations based on *tan* function in pigmentation biosynthesis and species pigmentation. Specifically, flies expressing the *D. americana tan* allele exhibited significantly darker pigmentation than those expressing the *D. novamexicana tan* allele in four of five experiments.

In addition to the full species alleles, chimeric alleles of *tan* were created that tested the effect of sequences in *tan* intron one, part of a previously mapped sequence implicated in pigmentation divergence between *D. americana* and *D. novamexicana* (Wittkopp et al., 2009). While the impact of these from these chimeric alleles was more nuanced, I found that noncoding sequence within intron 1 are sufficient to produce changes in pigmentation. This conclusion is supported by the results from a chimeric *tan* allele which replaces the 5' half of intron 1 sequence of *D. americana* with the corresponding sequence from *D. novamexicana* (A[N_5'_intron1]); this chimeric *tan* allele resulted in a significantly lighter pigmentation phenotype than the *D. americana* allele, providing evidence that noncoding changes in the 5' half of intron 1 contribute to pigmentation divergence.

The studies presented in this chapter provides an example of noncoding DNA sequence contributing to interspecies pigmentation divergence. While the work presented identified a smaller region of *tan* intron 1 than reported previously (Wittkopp et al., 2009), the ultimate goal would be to identify the causative nucleotide(s) in this region that lead to differences in pigmentation. With the advent of precise genome editing tools such as CRISPR, editing a single nucleotide to test its affect is becoming a feasible strategy. Future study in this system could

utilize CRISPR genome editing to test the effect of individual single nucleotide polymorphisms (SNPs), as well as combinations of SNPs, that exist in the 5' half of *tan* intron 1 between *D. americana* and *D. novamexicana.* Ideally, this type of single nucleotide replacement and testing would occur in the native species rather than the distantly related *D. melanogaster* or at least a more closely related host such as *D. virilis*. As CRISPR techniques in non-model species improve, this type of study will become more feasible. Identifying the causative nucleotide(s) that contribute to pigmentation divergence between these two species will add to a small, but growing number of cases where specific genetic changes in noncoding sequences are linked to phenotypic consequences. This type of information, especially when combined with information on the molecular mechanism through which specific noncoding sequence changes impact phenotypic divergence, will help in improving our ability to predict what impact, if any, changes in noncoding sequences will have on particular phenotypes.

Probing the mechanism through which non-coding sequences in tan *impact pigmentation divergence between* D. americana *and* D. novamexicana

With the implication of *tan* intron 1 in the phenotypic divergence between *D. americana* and *D. novamexicana,* I sought to identify the molecular mechanism through which this noncoding sequence influences pigmentation. Given the previous work showing a difference in transcriptional level gene expression of *tan* between *D. americana* and *D. novamexicana* and a role of *cis*-regulatory divergence in this, I hypothesized that noncoding sequence in intron 1 contains *cis*-regulatory elements that contribute to transcriptional level gene expression of *tan* and that have functional divergence between *D. americana* and *D. novamexicana*. I used three approaches to gain insight on this hypothesis: 1) Testing for differential *tan* mRNA expression in *tan* transgenic *D. melanogaster*, 2) Predicting the impact of sequence changes between species

on transcription factor binding sites (TFBS), and 3) Testing noncoding sequences in *tan* from *D. americana* and *D. novamexicana* for differences in capacity to activate gene expression, with green fluorescent protein (GFP) reporter genes.

Testing for differential *tan* mRNA expression in *D. melanogaster* expressing either *D. americana* or *D. novamexicana tan* failed to detect an expression difference in the direction expected from previous study (Cooley, Shefner, McLaughlin, Stewart, & Wittkopp, 2012). In the native species, *D. americana tan* allele is expressed more highly than the *D. novamexicana* allele. However, in the *D. melanogaster* transgenics, the *D. novamexicana tan* allele was found to be expressed at a significantly higher level than the *D. americana tan* allele. This result is especially surprising since the pigmentation phenotypes of the transgenic flies conforms to expectations based on pigmentation of the native species with the *D. americana tan* transgenics exhibiting a significantly darker pigmentation than their *D. novamexicana tan* counterparts. I postulate that the *D. melanogaster trans*-regulatory background may be sufficiently divergent from the native species and thus make investigating the molecular mechanism through which noncoding sequence changes act problematic. Future work examining the expression level driven by *D. americana* and *D. novamexicana tan* alleles and the chimeric alleles in a more relevant *trans*-regulatory background may provide better insight. For instance, the creation of these constructs, as well as *tan* alleles that change only single nucleotide that is divergent between species, in the native species or the more closely related *D. virilis*, could allow detection of subtle changes in *cis*-regulatory function that may be masked by the divergent *trans*-regulatory background of *D. melanogaster*.

If the noncoding changes in *tan* act to change mRNA level gene expression, I hypothesized they would do by changing (creating, destroying, or altering strength) of

transcription factor binding sites (TFBS) between species. Focusing on the 5' half of intron 1 implicated in pigmentation divergence between species by the *tan* transgenic analysis presented in Chapter 3, I found three transcription factors (TFs) with predicted binding sites surrounding the strongest candidate SNP between species: *AbdA, AbdB,* and *vvl*. Each of these has a predicted binding site in *D. americana* that is abolished (*AbdB, vvl*) or predicted reduced affinity (*AdbA*) in *D. novamexicana* and has a predicted impact on pigmentation that agrees with interspecies divergence (Rogers et al., 2014). A predicted binding site for *exd* exists in the sequence directly adjacent (and partially overlapping) these predicted binding sites; since *exd* is a cofactor that often works with *AbdA* and *AbdB* (Camino et al., 2015; Mann, Lelli, & Joshi, 2009; Slattery et al., 2011), the presence of its binding site gives further support to the hypothesis that these transcription factors may be regulating gene expression by binding to this sequence in *tan* intron 1. Future work to test for differential binding of these transcription factors to *D. americana* and *D. novamexicana* could provide further information about if these TFs and the predicted change in TFBS plays a role in pigmentation divergence between species.

Finally, the hypothesis that *tan* intron 1 contains an enhancer sequence was tested directly using a GFP reporter. Concurrently, the other large noncoding sequence in *tan,* intron 3, was also tested for enhancer function. Preliminary results suggest that all of these sequences (*D. americana* intron 1, *D. novamexicana* intron 1, *D. americana* intron 3, *D. novamexicana* intron 3) contain enhancer sequences. Future work will use these constructs to test for divergent function of these sequences between species by comparing the pattern and intensity of GFP expression. This work will give insight as to whether the noncoding sequences in *tan* intron 1 and *tan* intron 3 have differences in regulatory function between *D. americana* and *D. novamexicana.*

The genetic basis of pigmentation diversity within D. americana

The pigmentation diversity that exists within *D. americana* was used to study intraspecies phenotypic divergence. Specifically, I tested three main hypotheses: 1) *tan* and *ebony* contribute to pigmentation variation within *D. americana*, 2) additional genes contribute to intraspecies phenotypic diversity, and 3) genetic heterogeneity exists in *D. americana* such that different genetic loci produce phenotypically similar individuals. I generated F2 hybrid populations between different combinations of lightly and darkly pigmented *D. americana* lines and sorted the resulting individuals by pigmentation phenotype. The lightest and darkest individuals from each cross were pooled and genotyped at the pigmentation genes *tan, ebony,* and *yellow.* The results from this experiment support all three hypotheses. Future work should look for associations with additional loci on a genome-wide scale. Identification of the loci responsible for intraspecies pigmentation divergence in *D. americana* will allow for more conclusions about the similarities and differences of this divergence and the interspecies divergence between *D. americana* and *D. novamexicana*. Overall, this study offers a unique perspective into phenotypic evolution by allowing the same trait to be examined both within and between species. From the results presented here, this species system suggests that a combination of the same (*tan, ebony*) genes and different genes contribute to intraspecies and interspecies phenotypic divergence. In cases where *tan* and/or *ebony* are implicated, this system could be used to study the currently unanswered question regarding if the same or different nucleotides are responsible for phenotypic divergence.

References

Camino, E. M., Butts, J. C., Ordway, A., Vellky, J. E., Rebeiz, M., & Williams, T. M. (2015). The evolutionary origination and diversification of a dimorphic gene regulatory network

through parallel innovations in cis and trans. *PLoS Genet, 11*(4), e1005136. doi:10.1371/journal.pgen.1005136

Cooley, A. M., Shefner, L., McLaughlin, W. N., Stewart, E. E., & Wittkopp, P. J. (2012). The ontogeny of color: developmental origins of divergent pigmentation in Drosophila americana and D. novamexicana. *Evol Dev, 14*(4), 317-325. doi:10.1111/j.1525-142X.2012.00550.x

Mann, R. S., Lelli, K. M., & Joshi, R. (2009). Hox specificity unique roles for cofactors and collaborators. *Curr Top Dev Biol, 88*, 63-101. doi:10.1016/S0070-2153(09)88003-4

Rogers, W. A., Grover, S., Stringer, S. J., Parks, J., Rebeiz, M., & Williams, T. M. (2014). A survey of the trans-regulatory landscape for Drosophila melanogaster abdominal pigmentation. *Dev Biol, 385*(2), 417-432. doi:10.1016/j.ydbio.2013.11.013

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., . . . Mann, R. S. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell, 147*(6), 1270-1282. doi:10.1016/j.cell.2011.10.053

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution, 62*(9), 2155-2177. doi:10.1111/j.1558-5646.2008.00450.x

Sturtevant, A. H. (1925). The Effects of Unequal Crossing over at the Bar Locus in Drosophila. *Genetics, 10*(2), 117-147.

Wilson, C., Bellen, H. J., & Gehring, W. J. (1990). Position effects on eukaryotic gene expression. *Annu Rev Cell Biol, 6*, 679-714. doi:10.1146/annurev.cb.06.110190.003335

Wittkopp, P. J., Stewart, E. E., Arnold, L. L., Neidert, A. H., Haerum, B. K., Thompson, E. M., . . . Shefner, L. (2009). Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in Drosophila. *Science, 326*(5952), 540-544. doi:10.1126/science.1176980