Stony Brook University

The evolution of pigmentation traits in natural populations of *Drosophila melanogaster*

A Dissertation Presented

by

Rocio Siu Ng

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

December 2014

Stony Brook University

The Graduate School

Rocio Siu Ng

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

John R. True – Dissertation Advisor
Associate Professor, Department of Ecology and Evolution

Walter F. Eanes - Chairperson of Defense
Professor, Department of Ecology and Evolution

Douglas J. Futuyma
Distinguished Professor, Department of Ecology and Evolution

Paul Schmidt
Associate Professor, Department of Biology
University of Pennsylvania

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

The evolution of pigmentation traits in natural populations of *Drosophila melanogaster*

by

Rocio Siu Ng

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

2013

Research on insect melanism has been informative in understanding how evolutionary mechanisms can produce phenotypic diversity. My dissertation examines how genes contributing to the pigment biosynthesis pathway may be contributing to natural variation of pigmentation traits in Drosophila melanogaster. I uncovered significant variation in abdominal and thoracic pigmentation both within and among populations of D. melanogaster in the eastern United States. I closely examined how polymorphisms within the pigmentation genes, ebony, pale, Ddc, and tan are associated with pigmentation traits. I also looked for patterns of nucleotide diversity in each gene as well as evidence for selection. Each of the genes had loci that were significantly associated with pigmentation, with some being in regions that may be experiencing possible balancing or purifying selection. Adult expression of ebony and pale were also significantly correlated with pigmentation levels. These results suggest that various components in the pigmentation pathway are evolving together in order to produce phenotypic variation in these populations. Additionally, there was evidence for independent regulation of pigmentation expression in the thorax and the abdomen. Analysis of pigmentation phenotypes has revealed significant geographic patterns with a possible cline in thoracic traits. We also sequenced a group of alleles of the ebony and pale genes in these populations to examine clinal patterns and confirm associations with pigmentation traits. My work has the potential to increase the understanding of how polymorphisms at the nucleotide sequence level contribute to population level differences and possibly adaptation in D. melanogaster and other insect species.

Dedication Page

I would like to dedicate this dissertation to my family whom have been nothing but supportive throughout my career. Mom, Dad, and Celina, *Thank You* for always being there for me and being so understanding of the fact that I have been in school so long.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank first and foremost my advisor and mentor John True. From day one he has guided me through my career as a PhD Student and has provided me with endless support and advice. My dissertation committee, Walt Eanes, Doug Futuyma, and Paul Schmidt have been extremely helpful and have provided many insightful comments over the years. My undergraduate advisor David Fitch was a great inspiration in my decision to pursue gradate work and has been a great mentor and friend since.

Erik Lavington has been a great colleague who has made the transition into my new dissertation topic a lot easier and has never hesitated to help me or answer any questions. I cannot be more grateful that he was a member of the department.

I would also like to thank the following for general advice and help throughout my dissertation work: Shian-Ren Liou, Rodrigo Cogni, Abigail Cahill, Laurel Yohe, Spencer Koury, Shudan Yeh, Joseph Lachance, Matthew Aiello-Lammens, Paul Bourdeau, and Roman Yukilivech. Two of my undergraduate students Mark Lazare and Susan Ching were especially helpful in moving my data along and I have learned much from them. The Ecology and Evolution Department at Stony Brook University was a warm and rich place to complete my graduate work and I am grateful for the colleagues, faculty, and staff that I interacted with there.

I have also had the pleasure of meeting many beautiful people during my time at Stony Brook who had made my life there more rich and meaningful. I am extremely grateful for them and was incredibly lucky to have them all in my life. There are too many to list here so hopefully this paragraph suffices.

Special thanks goes to Patrick Lown who has been extremely supportive and has been with me through the best and the worst days of the last few years. He has also been instrumental in my development as a programmer which has helped me tremendously with my work. My friend Amy Pivak also needs to be acknowledged since she has done nothing but been an amazing friend and willing to venture far into Long Island just to hang out with me.

Lastly, I must acknowledge the staff at the Center for Inclusive Education, Nina Muang, Kathryne Piazzola, Tony Sperzal who have gone above and beyond to provide mentorship and guidance through this long process, and the Turner Fellowship Program and Turner family which have provided me with tremendous financial and emotional support and have been a great source of inspiration.

Chapter 1. Introduction

One of the longest standing pursuits of evolutionary biologists has been to understand the mechanisms that create and maintain phenotypic diversity. Many traits of interest are often quantitative in nature with multiple contributing genes of various effect sizes (Falconer and Mackay 1996). When studying these types of traits, scientists need to determine what loci are involved and what factors affect the segregation of contributing alleles in natural populations. It is also important to investigate how these alleles specifically influence gene expression and developmental processes that eventually lead to the final phenotype. For instance, a long standing question in evolution of development is whether coding or regulatory changes make up most of the genetic variation that contributes to phenotypic diversity. The role of natural selection on causative loci will also be informative in understanding how phenotypes evolve at the molecular level.

The evolution of pigmentation traits

Animal coloration is of particular interest since it is one of the most striking features of organisms, a clear demonstration of diversity, and is associated with many adaptations. Pigmentation traits thus have great potential to provide insight into how natural selection can promote diversity. While considerable levels of variation in color exist within and between species, it has been found to be involved in adaptations such as mimicry, aposematism, crypsis, and sexual ornamentation

(Cloudsley-Thompson 1999, Protas and Patel 2008). This suggests that selection affecting variation may be complex, since selection on one aspect of color may indirectly affect multiple traits. In *Drosophila,* for example, pigmentation has been connected to sexual selection (Yeh, Liou et al. 2006), desiccation tolerance (Kalmus 1941), and parasite resistance (Dombeck and Jaenike 2004). Each of these may have associated behaviors or physiological functions that may be affected by selection for high or low levels of pigmentation.

Insect pigmentation, in particular**,** has been extensively studied in many model species. Pigmentation in these species has been demonstrated to be highly variable, with incredible diversity at the population level in addition to diversity among species. Variation often exists in the types of pigments used, as well as their intensity and patterning. Relevant studies have aimed to understand the various functions of pigmentation, the evolution of pattern formation, and the underlying mechanisms governing the maintenance of variation (True 2003, Protas and Patel 2008, Wittkopp and Beldade 2009). There have been a variety of proposed evolutionary mechanisms at the molecular level to explain variation of pigmentation and patterning in various insect systems. For example, the origin of the derived trait of eyespot patterns in *Bicyclus anynana* has been shown to be a result of the formation of developmental organizers called foci in which signals to surrounding tissues using the protein Hedgehog (Hh), which also functions earlier in wing formation. Novel expression patterns of Hh and other patterning proteins such as Engrailed (En) and Cubitus interruptus (Ci) are all associated with the formation of the eyespot patterns in this

species (Brunetti, Selegue et al. 2001). In the *Heliconius* genus there is evidence for introgression of patterning alleles between populations of species that exhibit similar pigment patterns in sympatry (Martin, Dasmahapatra et al. 2013) .

Many studies have revealed the potential for pigmentation to be adaptive in many insect systems. As mentioned above, pigmentation has been associated with a number of traits such as body size, behavior, and immune functioning. Some of these links and diverse selective pressures associated with melanism have been studied in various insect species including those in the *Drosophila* genus. These selection pressures on pigmentation appear to be quite diverse among insect species (True 2003). There is a proposed role of melanin pigmentation in thermoregulatory function. As demonstrated in *Colias* butterflies, darker individuals were shown to absorb heat faster and retain a higher body temperature longer than lighter individuals (Watt 1968). In *Drosophila elegans*, black-morphs were on average .26°C higher than brown-morphs after experimental light irradiation (Hirai and Kimura 1997). Additionally, darker individuals of *Drosophila melanogaster* have been shown to have higher resistance to desiccation (Parkash, Kalra et al. 2010). In several species of insects, darker individuals were demonstrated to be more effective at resisting infection by pathogenic organisms. Some examples involve resistance to the parasitic nematode *Howardula aoronymphium* by *Drosophila falleni* (Dombeck and Jaenike 2004) and the entomopathogenic fungus, *Metarhizium anisopliae* by *Tenebrio molitor* (Barnes and Siva-Jothy 2000). Some studies have demonstrated the potential for pigmentation traits to be influenced by sexual selection. *Drosophila*

elegans males possess a melanic wing spot that is an integral component in courtship behavior (Yeh, Liou et al. 2006). *Drosophila melanogaster* males that are homozygous for the *ebony* mutation have been shown to have disadvantages in competitive mating (Kyriacou, Burnet et al. 1978). In *Pieris occidentalis* butterflies, marginal forewing melanization affects male mating success (Wiernasz 1989).

These studies suggest that expression of pigmentation and associated traits may be controlled by the same genes or by tightly linked genes. Some of these pleiotropic genes have been identified. For example, in several species of *Drosophila*, the *bric-á-brac2* (*bab2*) gene has been found to have pleiotropic functions in the regulation of pigmentation and trichome patterning (Gompel and Carroll 2003). Studying correlations of various traits with pigmentation has the potential to enable understanding of different selective pressures that may affect pigmentation. It will also be informative in understanding how selection can simultaneously affect a suite of traits.

Drosophila models

Species in the genus *Drosophila* have provided an excellent system for studying the evolution of pigment traits and patterning, since there are high levels of variation within and between species (Wittkopp, Carroll et al. 2003). In particular, most Drosophilidae have spatial patterns on their abdominal segments and/or thoraxes that vary in levels of light and dark melanization. (Wittkopp, Carroll et al. 2003). Many of these species are easily reared and manipulated in the laboratory

and have short generation times. Additionally, genetic tools available in *Drosophila melanogaster* have allowed researchers to target specific genetic and developmental pathways (Wittkopp and Beldade 2009). Many of the genetic and developmental mechanisms underlying pigmentation have become well understood (True 2003, Wittkopp, Williams et al. 2003, Vavricka, Christensen et al. 2010). Additionally, variation in pigmentation traits has been demonstrated in naturally occurring populations of *Drosophila melanogaster* throughout the world. This has been documented both in the thoracic trident, a pattern of pigmentation that forms a "trident-like" shape on the dorsal side of the thorax (David, Capy et al. 1985, Takahashi, Takahashi et al. 2007) and in abdominal stripe patterning (Kopp, Graze et al. 2003).

The demographic history of *Drosophila melanogaster* has provided a large "natural experiment" that can help us understand adaptation. The ancestral population of this species originated in Sub-Saharan Africa. *D. melanogaster* is thought to have undergone bottlenecks upon migration out of Africa, such that founder populations on different continents may possess distinct subsets of ancestral variation (David and Capy 1988, Baudry, Viginier et al. 2004). This species has subsequently faced novel temperate environments that are not present in their ancestral range. Studies have demonstrated that life history traits and the ability to diapause in *D. melanogaster* may have evolved in response to temperate environments and wintering conditions (Schmidt, Matzkin et al. 2005). Investigations of these traits and underlying influencing genes have also

demonstrated clinal patterns in North American populations reflecting adaptation like diapause are functionally involved in the ability of flies to overwinter (Schmidt and Paaby 2008, Schmidt, Zhu et al. 2008).

Pigmentation traits in *D. melanogaster* also exhibit both altitudinal and latitudinal clines. For example, altitudinal clines in abdominal pigmentation have been identified in populations in sub-Saharan Africa (Pool and Aquadro 2007). Similarly, latitudinal patterns have been identified in India highland and lowland populations (Munjal, Karan et al. 1997) and Australian populations (Telonis-Scott, Hoffmann et al. 2011). In all cases, increases in altitude or latitude are accompanied by increases in pigmentation levels in sample populations. These patterns suggest that pigmentation traits and underlying genes may potentially contribute to adaptation. Gene flow between neighboring populations is one possible mechanism acting to maintain variation in individual populations.

Altitudinal but not clinal patterns were found in Sub-Saharan Africa, which is representative of the ancestral population of *Drosophila melanogaster*. These may reflect adaptation to similar climatic conditions associated with higher altitudes and increasing latitudes in temperate environments. However, because *D. melanogaster* has been demonstrated to have undergone bottlenecks (Baudry, Viginier et al. 2004), these patterns elsewhere may reflect convergent evolution in pigmentation traits and genes in response to novel temperate climates. Desiccation tolerance, which is positively associated with pigmentation, also increases with higher latitudes (Parkash, Rajpurohit et al. 2008). Clinal patterns are present in thermal sensitivity

in *D. melanogaster* (Hoffmann, Anderson et al. 2002) and pigmentation has been proposed to regulate body temperature in other insects (Watt 1968, Hirai and Kimura 1997) These studies suggest adaptation to temperature and climatic gradients may be driving observed clinal patterns in pigmentation levels.

Pigmentation biosynthesis pathway

In insects, pigment molecules typically are synthesized in the epidermal cells and incorporated into the exoskeleton through the process of sclerotization (Wright 1987). Cuticular sclerotization, which can occur before or soon after ecdysis, stabilizes the insect cuticle through the incorporation of cross-linked and polymerized phenolic compounds (Andersen 2010). The distribution of the pigments is mediated by a set of patterning genes that regulate expression of effector genes that encode enzymes involved in the pigment biosynthesis pathway (Wittkopp and Beldade 2009). In *D. melanogaster* this process is well understood and many of the genes have been characterized. Pigment patterning is controlled by regulatory proteins that have pleiotropic effects in developmental processes, such as HOX genes, *optomotor-blind*, *bric-a-brac*, and *engrailed* as well as sex-determination genes (Wittkopp, Carroll et al. 2003).

Melanin pigments are synthesized by a biochemical pathway that converts tyrosine into Dopa and dopamine precursors through the enzymatic activity of Tyrosine-Hydrozylase (*pale*) and Dopa Decarboxylase (*Ddc*). The pathway branches into various reactions that produce brown, black, and yellowish-tan pigments that

are then deposited into the sclerotizing cuticle (Figure 1.1, reviewed in True 2003). The spatiotemporal regulation of the *yellow*, *tan* and *ebony* effector genes determine the distribution and abundance of these pigments (Wittkopp, True et al. 2002). For example, *yellow* mutants lack dark coloration, demonstrating that *yellow* is needed for the production of black melanin (Morgan and Bridges 1916). Expression of *ebony* along with *yellow* is required for the formation of abdominal stripes in *D. melanogaster*. This suggests that expression of these two genes interact to produce specific melanic patterns and may contribute to species-specific coloration (Wittkopp, True et al. 2002).

While genes involved in melanin production and patterning have been well studied, some have been implicated in within- and between-species diversity. *Yellow, tan, pale* and *ebony*, in particular have been well documented in cases of intra- and interspecific variation in *Drosophila* species (Table 1.1). Strong haplotype structure in noncoding regions of *ebony* has been associated with altitudinal increases in pigmentation in Sub–Saharan populations of *D. melanogaster* (Rebeiz, Pool et al. 2009). A genome-wide association study (GWAS) study in European populations of *D. melanogaster* has also associated regulatory regions in *tan* with variation in abdominal pigmentation (Bastide, Betancourt et al. 2013). Regulatory elements in *ebony* and *tan* have been also been associated with divergence in body color between *D. novamexicana* and *D. Americana* (Wittkopp, Stewart et al. 2009) and *tan* has been implicated in abdominal pigmentation differences between *D. yakuba* and *D. santomea* (Jeong, Rebeiz et al. 2008).

Outline of Dissertation

My dissertation addresses how pigmentation pathway genes may influence variation and the evolution of pigmentation traits in *Drosophila melanogaster*. I utilized both inbred and extraction lines of *Drosophila melanogaster* derived from natural populations of the eastern seaboard of the United States. The next chapter describes work that involved isogenic lines derived from a single population as part of the Drosophila Genetic Reference Panel (DGRP). I used phenotypic data I collected in conjunction with genomic sequence and expression data available from the panel. Through a SNP association study, I have identified regulatory regions and coding changes in the genes *ebony*, *pale*, *ddc*, and *tan* that may be responsible for differences in phenotypic expression. I have also conducted polymorphism analyses each of these genes to determine how these causative alleles may be evolving in this population.

The next chapter expands the results from the DGRP lines, by discussing results obtained from the phenotypic and genetic analysis of third chromosome extraction lines. Extraction lines place chromosomes of interest on common genetic backgrounds, in order to partially control for epistatic effects. The third chromosome contains two of the previously studied candidate genes, *ebony* and *pale*. The goal of this study was to determine how these genes, and possibly other genetic factors on the third chromosome may contribute to variation in pigmentation traits within and among populations in the Easter United States. The wild derived strains used to generate the extraction lines were collected by W. Eanes and colleagues and represent

populations from Maine, Pennsylvania, North Carolina, Northern Florida and Southern Florida. Subsequent sequencing investigated how these candidate genes may have evolved in these populations and whether they may be subject to selection. My work has the potential to increase the understanding of how polymorphisms at the nucleotide sequence level contribute to population level differences, geographical variation, and adaptation in *D. melanogaster* and other insect species.

References

Andersen, S. O. (2010). "Insect cuticular sclerotization: A review." Insect Biochemistry and Molecular Biology **40**(3): 166-178.

Barnes, A. I. and M. T. Siva-Jothy (2000). "Density-dependent prophylaxis in the mealworm beetle *Tenebrio molitor* L. (Coleoptera: Tenebrionidae): cuticular melanization is an indicator of investment in immunity." Proc Biol Sci **267**(1439): 177-182.

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

Baudry, E., B. Viginier and M. Veuille (2004). "Non-African populations of *Drosophila melanogaster* have a unique origin." Molecular Biology and Evolution **21**(8): 1482-1491.

Brunetti, C. R., J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield and S. B. Carroll (2001). "The generation and diversification of butterfly eyespot color patterns." Current Biology **11**(20): 1578-1585.

Cloudsley-Thompson, J. L. (1999). "Multiple factors in the evolution of animal coloration." Naturwissenschaften **86**(3): 123-132.

David, J. R. and P. Capy (1988). "Genetic variation of *Drosophila melanogaster* natural populations." Trends in Genetics **4**(4): 106-111.

David, J. R., P. Capy, P. Véronique and S. Tsakas (1985). "Thoracic trident pigmentation in *Drosophila melanogaster*: Differentiation of geographical populations." Genetics Selection Evolution **17**(2): 211-224.

Dombeck, I. and J. Jaenike (2004). "Ecological genetics of abdominal pigmentation in *Drosophila falleni*: A pleiotropic link to nematode parasitism." Evolution **58**(3): 587-596.

Falconer, D. S. and T. F. C. Mackay (1996). Introduction to quantitative genetics. Essex, England, Longman.

Gompel, N. and S. B. Carroll (2003). "Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies." Nature **424**(6951): 931-935.

Hirai, Y. and M. T. Kimura (1997). "Incipient reproductive isolation between two morphs of *Drosophila elegans* (Diptera: Drosophilidae)." Biological Journal of the Linnean Society **61**(4): 501-513.

Hoffmann, A. A., A. Anderson and R. Hallas (2002). "Opposing clines for high and low temperature resistance in *Drosophila melanogaster*." Ecology Letters **5**(5): 614-618.

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

Kalmus, H. (1941). "The resistance to desiccation of *Drosophila* mutants affecting body colour." Proceedings of the Royal Society of London Series B-Biological Sciences **130**(859): 185-201.

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

Kyriacou, C. P., B. Burnet and K. Connolly (1978). "The behavioural basis of overdominance in competitive mating success at the *ebony* locus of *Drosophila melanogaster*." Animal Behaviour **26**(Part 4): 1195-1206.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet and C. D. Jiggins (2013). "Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies." Genome Research **23**(11): 1817-1828.

Morgan, T. H. and C. B. Bridges (1916). Sex-linked inheritance in *Drosophila*. Carnegie Institution of Washington publication 237. Washington,, Carnegie Institution of Washington,.

Munjal, A. K., D. Karan, P. Gibert, B. Moreteau, R. Parkash and J. R. David (1997). "Thoracic trident pigmentation in *Drosophila melanogaster*: latitudinal and altitudinal clines in Indian populations." Genetics Selection Evolution **29**(6): 601-610.

Parkash, R., B. Kalra and V. Sharma (2010). "Impact of body melanisation on contrasting levels of desiccation resistance in a circumtropical and a generalist *Drosophila* species." Evolutionary Ecology **24**(1): 207-225.

Parkash, R., S. Rajpurohit and S. Ramniwas (2008). "Changes in body melanisation and desiccation resistance in highland vs. lowland populations of *Drosophila*. melanogaster." J. Insect Physiol. Journal of Insect Physiology **54**(6): 1050-1056.

Pool, J. E. and C. F. Aquadro (2007). "The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*." Molecular Ecology **16**(14): 2844-2851.

Protas, M. E. and N. H. Patel (2008). "Evolution of Coloration Patterns." Annual Review of Cell and Developmental Biology **24**: 425-446.

Rebeiz, M., J. E. Pool, V. A. Kassner, C. F. Aquadro and S. B. Carroll (2009). "Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population." Science **326**(5960): 1663-1667.

Schmidt, P. S., L. Matzkin, M. Ippolito and W. F. Eanes (2005). "Geographic variation in diapause incident, life-history traits and climatic adaptation in *Drosophila melanogaster*." Evolution **59**(8): 1721-1732.

Schmidt, P. S. and A. B. Paaby (2008). "Reproductive diapause and life-history clines in North American populations of *Drosophila melanogaster*." Evolution **62**(5): 1204-1215.

Schmidt, P. S., C.-T. Zhu, J. Das, M. Batavia, L. Yang and W. F. Eanes (2008). "An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*." Proceedings of the National Academy of Sciences **105**(42): 16207-16211.

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

Telonis-Scott, M., A. A. Hoffmann and C. M. Sgrò (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.

True, J. R. (2003). "Insect melanism: the molecules matter." Trends in Ecology & Evolution **18**(12): 640-647.

Vavricka, C., B. Christensen and J. Li (2010). "Melanization in living organisms: a perspective of species evolution." Protein & Cell **1**(9): 830-841.

Watt, W. B. (1968). "Adaptive significance of pigment polymorphisms in *Colias* butterflies. I. Variation of melanin pigment in relation to thermoregulation." Evolution **22**(3): 437-458.

Wiernasz, D. C. (1989). "Female choice and sexual selection of male wing melanin pattern in *Pieris occidentalis* (Lepidoptera)." Evolution **43**(8): 1672-1682.

Wittkopp, P. J. and P. Beldade (2009). "Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy." Seminars in Cell & Developmental Biology **20**(1): 65-71.

Wittkopp, P. J., S. B. Carroll and A. Kopp (2003). "Evolution in black and white: genetic control of pigment patterns in *Drosophila*." Trends in Genetics **19**(9): 495-504.

Wittkopp, P. J., E. E. Stewart, L. L. Arnold, A. H. Neidert, B. K. Haerum, E. M. Thompson, S. Akhras, G. Smith-Winberry and L. Shefner (2009). "Intraspecific polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*." Science **326**(5952): 540-544.

Wittkopp, P. J., J. R. True and S. B. Carroll (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., B. L. Williams, J. E. Selegue and S. B. Carroll (2003). "*Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes." Proceedings of the National Academy of Sciences of the United States of America **100**(4): 1808-1813.

Wright, T. R. (1987). "The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*." Adv. Genet **24**(2): 127.

Yeh, S. D., S. R. Liou and J. R. True (2006). "Genetics of divergence in male wing pigmentation and courtship behavior between *Drosophila elegans* and *Drosophila gunungcola*." Heredity **96**(5): 383-395.

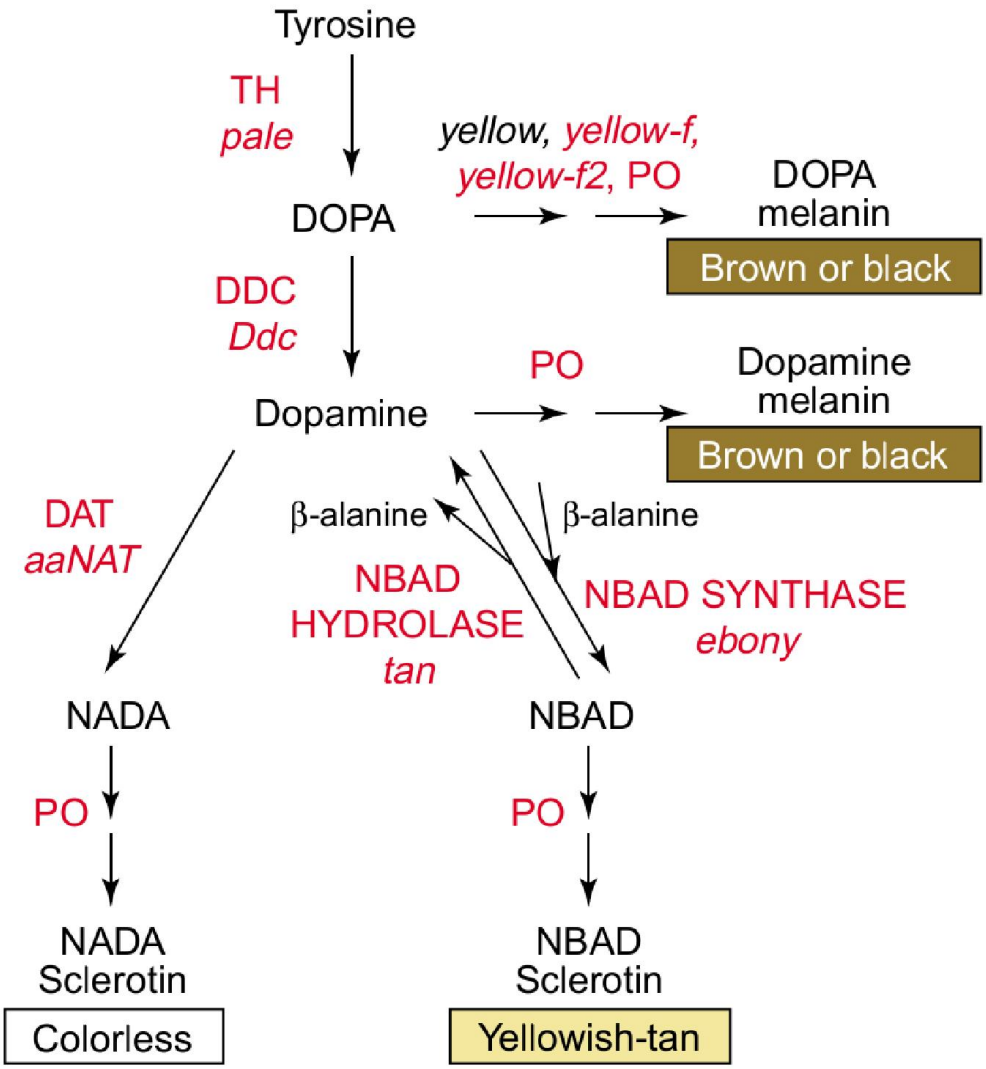


Figure 1.1. Current model of the biosynthesis pathway and contribution candidate genes for melanin pigment production. Red text indicates acting genes (italicized) and encoded enzymes (capitalized). Final pigments are at the ends of each pathway route with colored boxes representing pigment types. Abbreviations: DAT (dopamine acetyltransferase); DDC (DOPA decarboxylase); DOPA (dihydroxyphenylalanine); NADA (N-acetyl dopamine); NBAD(N-b-alanyl dopamine); PO, (phenoloxidases); TH(tyrosine hydroxylase).

Table 1.1. An overview of the candidate pigmentation examined in this study and known roles in species divergence and intraspecific variation in studies with Drosophila

Gene (Protein)	Role in interspecific variation or species divergence in pigmentation traits	Proposed Nature of genetic variation	References
ebony (NBAD Synthase)	Variation in abdominal pigmentation in African populations of *D. melanogaster*	regulatory	(Pool and Aquadro 2007), (Rebeiz, Pool et al. 2009)
	Variation in thoracic pigmentation in two wild derived strains of *D. melanogaster*	coding	(Takahashi, Takahashi et al. 2007)
	Clinal variation in thoracic trident pigmentation in east coast Australian populations of *D. melanogaster*	regulatory	(Telonis-Scott, Hoffmann et al. 2011)
	Interspecific variation in overall body color United States populations of *D. Americana*	regulatory	(Wittkopp, Stewart et al. 2009)
	Divergence in overall body color between *D. novamexicana* and *D. americana*	regulatory	(Wittkopp, Williams et al. 2003), (Wittkopp, Stewart et al. 2009)
pale (Tyrosine Hydroxylase)	Possible contribution to *Drosophila* wing melanin patterning differences		(True, Edwards et al. 1999)
yellow	Divergence in body pigmentation in *D. melanogaster*, *D. subobscura* , and *D. virilis*	regulatory	(Wittkopp, Vaccaro et al. 2002)
tan (NBAD Hydrolase)	Variation in abdominal pigmentation in European populations of *D. melanogaster*	regulatory	(Bastide, Betancourt et al. 2013)
	Variation in overall body color in United States populations of *D. Americana*	regulatory	(Wittkopp, Stewart et al. 2009)
	Divergence in overall body color between *D. novamexicana* and *D. americana*	regulatory	(Wittkopp, Stewart et al. 2009)

Chapter 2. The evolutionary genetics of phenotypic variation in pigmentation traits in *Drosophila melanogaster*

Introduction

The study of insect pigmentation has been informative for studying the mechanisms underlying phenotypic evolution. There is considerable diversity in pigmentation traits, which have been associated with various functions such as thermoregulation (Watt 1968), behavior (Yeh, Liou et al. 2006), desiccation tolerance (Parkash, Kalra et al. 2010), and immunity (Barnes and Siva-Jothy 2000). This suggests that pigmentation may be subject to complex selection pressures. The simplest form of pigmentation variation, melanism of the body, is present in many animal species (Majerus 1998). There are many documented examples of pigmentation variation, from wing patterning differences in *Heliconius* butterflies (Joron, Jiggins et al. 2006) and melanic forms of squamate lizard species (Rosenblum, Hoekstra et al. 2004) to variation in coat coloration in populations of rock pocket mice (Hoekstra, Krenz et al. 2004).

Insect models such as *Drosophila* species have been useful in studying the evolution of pigmentation traits and underlying genetic mechanisms. There is significant inter- and intraspecific variation in melanin patterning among *Drosophila* species (Wittkopp, Carroll et al. 2003). In some cases, divergence in pigmentation traits between closely related species has been demonstrated to correlate with

divergence in gene expression. For example, melanin patterning differences in *D. melanogaster*, *D. subobscura*, and *D. virilis* have been causally linked to divergence in regulation of expression of YELLOW, a cuticular protein involved with processing of the melanin precursor DOPA (dihydroxyphenylalanine) (Wittkopp, Vaccaro et al. 2002). It would be interesting to determine what population level processes, such as natural and sexual selection, underlie pigmentation diversity. Some melanin patterns in *Drosophila* are sex specific and may be involved in sexual selection (Kopp, Duncan et al. 2000, Kopp and True 2002).

The biochemical pathway leading to the production and deposition of melanin pigments from precursor molecules is highly conserved and contains several well-characterized genes (Wittkopp, Carroll et al. 2003, Futahashi, Sato et al. 2008, Yu, Shen et al. 2011). Insect epidermal cells produce melanin and related pigments using a set of core enzymatic reactions encoded by the pigmentation genes *pale* (Tyrosine Hydroxylase), *Ddc* (Dopa Decarboxylase), *ebony* (N-ß-alanyl Dopamine Synthase), and *tan* (N-ß-alanyl Dopamine Hydrolase). From this pathway, four different pigment precursors are secreted into the developing cuticle (Figure 2.1). Accumulation of these pigments results in the overall cuticular color and melanin patterns, such as abdominal striping and the thoracic trident, (Wright 1987, True 2003).

The presence of a well characterized pathway provides us with the unique opportunity to explore how different components may act together in order to produce observed levels of natural variation. There is a growing body of evidence that suggests that the positions of enzymes within metabolic pathways may influence

their evolutionary trajectories (Flowers, Sezgin et al. 2007). For instance, models of enzyme kinetics have predicted that flux control within a linear pathway will evolve to be unequal among enzymes in the pathway and may be related to the level of adaptive substitutions found in these enzymes. During adaptive evolution, upstream enzymes are expected to fix most of the advantageous mutations and evolve higher control coefficients than downstream enzymes. Near the fitness optimum, this trend is predicted to reverse with increased substitutions in downstream enzymes due to relaxed selection in these enzymes and strong purifying selection on upstream enzymes (Wright and Rausher 2010). Empirical work has supported some of these predictions by uncovering elevated rates of adaptive evolution of upstream enzymes in metabolic pathways in *Drosophila melanogaster* (Flowers, Sezgin et al. 2007), and the Glucosinolate pathway in *Arabidopsis thaliana* (Olson-Manning, Lee et al. 2013). Higher rates of substitution in downstream genes have been demonstrated in the Anthocyanin Pathway (Lu and Rausher 2003) , while lower genetic diversity was observed in upstream genes of the melanin synthesis pathway in silkworms (Yu, Shen et al. 2011) which may have been due to relaxed constraints on these genes as predicted by the model of Wright and Rausher (2010). These studies demonstrate the importance of considering the positions of enzymes within metabolic pathways and how this may effect genetic polymorphisms and impact phenotypes of interest.

Since several loci involved in *Drosophila* melanin biosynthesis have been well characterized, this allows for the implementation of a candidate gene approach to investigate the genetic basis of population level variation in pigmentation. The

genetic and developmental mechanisms responsible for variation and evolution of pigmentation traits remain to be fully understood. The aim of this work is to lead to a better understanding of how specific pigmentation genes within the biosynthesis pathway are evolving and may be contributing to natural variation in pigmentation traits in *Drosophila melanogaster*. Variation has been observed within and among naturally occurring populations of *Drosophila melanogaster* throughout the world. This has been documented in the thoracic trident, a pattern of pigmentation that forms a "trident-like" shape on the dorsal thorax (David, Capy et al. 1985, Takahashi, Takahashi et al. 2007) and abdominal stripe patterning (Kopp, Graze et al. 2003). In addition to identifying responsible genes, we also hope to determine the nature of genetic variation contributing to variation, i.e. whether causative loci are in coding or regulatory regions.

This study focuses on candidate genes (*ebony*, *pale*, *tan* and *Ddc*) that are major players in the biosynthesis pathway that produces melanin and related molecules (True 2003, Wittkopp and Beldade 2009). Some of these genes have been associated with species divergence as well as intraspecific variation in pigmentation traits in *Drosophila* (Table 2.1). The *ebony* locus, which encodes N-ß-alanyl dopamine synthetase and is located on the third chromosome, is of particular interest since it has been implicated in variation in abdominal stripe patterning (Rebeiz, Pool et al. 2009) and thoracic pigmentation variation (Takahashi and Takano-Shimizu 2011) among natural populations of *D. melanogaster*. *ebony* expression differences have also been implicated in divergence in body color between *D. novamexicana* and *D.*

americana (Wittkopp, Williams et al. 2003). The expression of *ebony* has also been demonstrated to work complementarily with the *yellow* gene in determining the patterning and intensity of pigment production (Wittkopp, True et al. 2002). *Ddc*, which encodes dopa decarboxylase, and *pale*, which encodes tyrosine-hydroxylase (Neckameyer and White 1993) are also of interest because they encode products involved the early steps of the melanin biosynthesis pathway (Figure 2.1, True 2003). They are located on the second and third chromosomes respectively.

This study makes use of the isogenic lines developed as part of the Drosophila Genetics Reference Panel (DGRP) project to identify potential candidate sequence variants in pigmentation genes that may contribute to variation in thoracic and abdominal pigmentation. The panel consists of 192 sequenced inbred lines that were derived from a single population from Raleigh, North Carolina (Mackay, Richards et al. 2012). Pigmentation phenotype assays were completed on 32 of these lines and available DGRP sequence data were used to carry out a SNP association study. Phenotypic data were used in conjunction with available DGRP expression data (Ayroles, Carbone et al. 2009, Massouras, Waszak et al. 2012) to determine if adult gene expression is associated with pigmentation differences.

Materials and Methods

Flies used from this study were from lines developed as part the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay, Richards et al. 2012). These lines are representative of a single population from Raleigh, North Carolina, which were inbred by 20 generations of full-sib mating to make them effectively isogenic (Ayroles, Carbone et al. 2009). One hundred and sixty-eight of these lines were sequenced using both Illumina and 454 sequencing technology (Mackay, Richards et al. 2012). Additionally RNA expression profiles of 3- to 5-day old flies were completed and available from 40 of the DGRP lines (Ayroles, Carbone et al. 2009, Massouras, Waszak et al. 2012). I assayed pigmentation levels in 32 of the DGRP lines that had both sequence and expression data available. Pigmentation was measured on the thorax and abdomen of adult female flies about a week post-eclosion (see below).

Phenotyping assays

The DGRP lines were obtained from the Bloomington Drosophila Stock Center (http://flystocks.bio.indiana.edu/). Five cultures of five males and five females from each line were set up on standard molasses/corn meal/yeast extract medium at 21˚C. After four to five days the parents were removed. Newly emerging female adults were collected from the three most prolific crosses, and aged in vials for five to seven days. Five flies per vial (15 per line) were individually photographed under 5X magnification using an AxioCam (Zeiss) video camera attached to a Leica MZ7 microscope. AxioVision software (Rel. 4.3) was used to capture the images using uniform settings and lighting. Images included specific regions on thoracic and

abdominal segments. Measured regions included those within and outside the thoracic trident, and within and outside the stripe on the A4 abdominal segment Figure 2.2, Figure 2.2). These regions were chosen based on consistent visibility and absence of glares and bristle obstruction. Luminosity, which measures the brightness of selected pixels in an image, was recorded in areas of interest using Image J software (Schneider, Rasband et al. 2012). Luminosity of black and yellow beads included in the images were measured to standardize for random changes in lighting. Pigmentation luminosity was standardized as a number between 0 (black bead luminosity) and 1 (yellow bead luminosity). A Nested Analysis of Variance (ANOVA) was used to analyze line differences in all assayed pigmentation traits. Phenotypic correlations were estimated by calculating correlation coefficients among pairs of pigmentation traits between line averages of luminosity measurements. The Pearson product moment correlation coefficient was estimated to test for the strength of the linear relationship between the traits and to calculate its statistical significance. Genetic correlations were calculated for each trait pair using methods described by (Hegmann and Possidente 1981) for estimating genetic correlations on inbred lines. These methods uses among strain components of covariance to estimate additive genetic covariances and thus genetic correlations. Statistical analyses were carried out using R software (R Core Team 2013).

SNP Association

Sequence data from the candidate genes *ebony*, *pale*, *Ddc,* and *tan* were extracted from the original 37 sequences of the DGRP, which contained the lines that

were phenotyped for pigmentation (Mackay, Richards et al. 2012). All exons and introns for each gene were included as well as the 5 kb upstream of the start codon. Sequences obtained from Flybase were used as reference sequences to identify each gene sequence in the DGRP data (Marygold, Leyland et al. 2013). MEGA 6 (Tamura, Stecher et al. 2013) was used to align the sequences to the reference and isolate polymorphic sites among the lines. Sites with too many non-reads (more than %50) or low allele frequencies (less than 3 affected lines) were filtered out manually to insure the quality of the tested sites.

For each included polymorphic site, a Nested Analysis of Variance (ANOVA) was used to test for associations between alternative SNPs and differences in the four measured pigmentation traits among the 32 phenotyped lines. The effects of each SNP allele, nested lines in SNP alleles, and nested vials within each line were estimated. Singletons were excluded from this analysis since any causative effects are unlikely to be detected with the current sample size. To resolve the issue of multiple testing, 10,000 random permutations of the phenotypic data were carried out across lines while keeping allele states intact, similar to the method used in Lavington, Cogni et al. (2014). Each permutation generated separate nested ANOVA calculations and distributions of F-ratios. SNP loci with *F*-values higher than the 95% percentile of the distribution were considered significant. A separate permutation test was conducted for each phenotypic trait (pigmentation and expression) and gene across all SNPS within the genic and upstream regions. This test was performed using R Software (R Core Team 2013). In order to determine the

magnitude of allele effect sizes, luminosity scores were standardized by calculating the number of standard deviations each observation was above or below the mean. These were calculated for associated traits of significant SNP sites for each allele state.

Expression Analysis

Whole-adult Affymetrix expression data reported by Aryoles et al. (2009) were downloaded and probes with underlying SNPS were removed or masked since these can affect expression array performance (Benovoy, Kwan et al. 2008, Chen, Page et al. 2009). Only data from female flies were included since pigmentation was only measured in females in this study. A SNP association study was also carried out on this data using the permutation methods outlined above to see if any SNP sites were associated with expression. A multivariate linear regression was carried out to determine if mRNA expression of the pigmentation genes *ebony*, *pale*, *tan* and *Dopa decarboxylase (Ddc)* were correlated with pigmentation levels.

Sequence Analysis

The DNA sequences for each gene were extracted from the DGRP database and filtered and modified based on accompanying quality scores. Non-reads and sites that were below a Phred score of 30 (99.9% base call accuracy) were converted to the majority allele (Table 2.5). For each candidate gene, DNAsp(Librado and Rozas 2009) (Ayroles, Carbone et al. 2009) was used to calculate various polymorphism parameters, including the number of segregating sites (*S*), Watterson's Theta

estimator ($\underline{\theta}_w$) (Watterson 1975), and nucleotide diversity (π) (Tajima 1983). These parameters were calculated for the whole gene sequence (exons and introns), and the upstream 5kb region. The diversity estimates of $\underline{\theta}_w$ and π were compared to chromosome level estimates derived for the DGRP populations, which included the lines in the present study (Mackay, Richards et al. 2012). Sliding window analyses of nucleotide diversity (π) were completed on each gene and upstream region inclusively. For the entire gene region (exons and introns) and upstream regions of *ebony*, *pale*, *ddc*, and *tan*, window lengths of 500 bp were used with overlapping steps of 25 bp. Since the sequence length for *Ddc* was considerably shorter than the rest, window lengths of 400 bp were used with step sizes of 20 bp. Linkage disequilibrium measures (*D'* and R^2) were calculated among polymorphic sites within the gene and upstream regions, using Fisher's exact test and X^2 estimates to test for significance. *D'* and R^2 values were plotted against pairwise distance for the significant SNP sites within each gene.

Neutrality Tests

DNAsp (Librado and Rozas 2009) was used to calculate the Jukes-Cantor corrected divergence rates using sequence data from *Drosophila simulans*. *D. simulans* sequence data were obtained from Flybase (Marygold, Leyland et al. 2013) and White105 sequence data available from the 12 Drosophila genomes Consortium 2007 (version r1.01 released 2014). Alignments were performed using MUSCLE (Edgar 2004) algorithms in the software MEGA 6 (Tamura, Stecher et al. 2013). To examine any deviations of the gene sequences from neutral evolution, Tajima's D test

(Tajima 1989), Fay and Wu's H test (Fay and Wu 2000), E test (Zeng, Fu et al. 2006), the Mcdonald-Kreitman (MK) test (McDonald and Kreitman 1991), and the HKA test (Hudson, Kreitman et al. 1987) were all performed using sequence data from the DGRP lines and *D. simulans*.

For each candidate gene sequence from the DGRP, Tajima's *D* was calculated separately on the gene (exons and introns) and 5kb upstream sequences from the DGRP lines using DNAsp (Librado and Rozas 2009). Tajima's *D* was calculated for all sites based on the discrepancy between the nucleotide diversity per site (π) and Watterson's theta estimator (θ_w). Sliding window analysis of Tajima's *D* was also completed in DNAsp along the gene and upstream regions inclusively using the same parameters for π (see above). Fay and Wu's *H* (Fay and Wu 2000) and Zeng et al.'s *E* (Zeng, Fu et al. 2006) was calculated for the coding regions (CDS) and the gene regions of each candidate gene using the *D. simulans* sequence as the outgroup. *H* measures departures from neutrality reflected in differences between high frequency and intermediate alleles and is less sensitive to population expansion than Tajima's *D*. *E* contrasts estimates of θ derived between high and low frequency alleles and can be used to detect selective sweeps. These tests were performed using the DH program (http://zeng-lab.group.shef.ac.uk/wordpress/?page_id=28), which uses models as described in (Zeng, Fu et al. 2006). Significance levels were determined by carrying out 10,000 coalescent simulations with the number of segregating sites fixed. Tajima's *D* (Tajima 1989) was also calculated for the CDS and gene regions using the same methods.

MK tests were carried out on coding regions for each gene using the DGRP sequence data (N=37) and the *D. simulans* sequence (N = 1). This test compares the ratio of non-synonymous to synonymous variation within and between species. Fisher's exact test and a *G* test were used to determine statistical significance and the *G* value was modified by Williams' correction (Cochran 1954). The HKA test (Hudson, Kreitman et al. 1987), which is a multi-locus test, was used to compare polymorphism and divergence estimates among sequence data for the four candidate genes (*ebony*, *pale*, *Ddc*, and *tan*) in the DGRP lines, using the consensus sequence from *D. simulans* as an outgroup. This test was conducted using the HKA program (https://bio.cst.temple.edu/~hey/software/software.htm) which uses coalescent simulations to estimate population parameters and expected values of θ under neutral expectations (Hey 2001).This test was run for 10,000 simulations.

Results

Thirty-two isogenic lines from the Drosophila Genetics Reference Panel (DGRP) were assayed to access variation in female pigmentation traits in a single population from Raleigh, NC. Luminosity measurements, which indicate the brightness of a sampled segment, were made for trident and non-trident areas of the thorax and stripe and non-stripe areas of abdominal segment 4 (Figure 2.2).

Phenotypic Variation

There was considerable variation in both thoracic and abdominal pigmentation traits among the 32 lines that were assayed for these phenotypes (Figure 2.3, Figure 2.4). A nested ANOVA analysis revealed significant levels of variation among the lines (F values ranged from 7.70 to 15.46, $P < 0.0001$) using luminosity measurements in all the measured pigmentation traits (Table 2.2). Strong correlations were observed between all pairwise combinations of the four traits (Correlation coefficients ranged from 0.33 to 0.85, $P <.0001$). However, correlation coefficients were higher for pigmentation values representative of the same tagma (0.80 to 0.85) than those on different tagmata (0.33 to 0.75) of the adult fly (Table 2.3). Genetic correlations were also estimated for each trait pair. Again, estimates were highest for pigmentation traits on the same tagma (0.40 to 0.41) and lowest between the trident and abdominal striping (0.17, Table 2.4).

SNP Associations

A SNP association study was conducted to determine if any of the polymorphic sites in the gene region (exons and intron) and upstream regions of *ebony*, *pale*, *Ddc*,

and *tan* were associated with pigmentation variation and mRNA expression (Table 2.7, Table 2.10). Due to the nature of the DGRP sequence data, only polymorphic variation was identifiable, since indels and repeats were excluded from the sequence data. The entire genic region, as well as 5kb region upstream of the start codon, was included in this analysis. Numbered positions referred to where the SNP sites resides within the tested gene region starting at the end position of each gene and ending 5kb upstream of the gene. There were significant SNPs in all the investigated genes, with a vast majority located in non-coding regions (Table 2.6). The significantly associated SNPs also varied among the pigmentation traits. For example, SNP sites that were associated with thoracic traits were often not associated with abdominal traits, and vice versa (Table 2.7, Figure 2.5). There were also several significant SNP sites associated with expression levels in each candidate gene, with the *pale* gene containing the largest number of significant SNPs (Table 2.10). The two sites in *ebony* associated with expression were also found to be associated with pigmentation traits. None of the significant sites for *pale*, *Ddc*, or *tan* expression were also associated with pigmentation.

Effect sizes varied among the alleles within each gene and among the genes (Table 2.8, Table 2.9). The average negative effect sizes were generally larger in *ebony*, *Ddc* and *tan*, while in *pale* the positive effects were larger. Negative effect sizes reflect a negative effect on luminosity measurements, which means that there is an increase in pigmentation intensity. In other words polymorphisms in *pale* overall tend to decrease pigmentation while polymorphisms in the other genes have

an overall positive effect on pigmentation intensity. No single gene had the largest positive or negative effect size for all the traits.

Sequence diversity of candidate genes

Measures of nucleotide diversity (π and θ_w) were estimated for each gene (introns and exons) and its 5kb upstream region and compared to estimates for chromosome arm in the DGRP lines from Mackay, Richards et al. 2012. (Table 2.11). In most cases π and θ_w were higher for the upstream regions than the gene regions, with the exception of θ_w estimates in *pale*. Most estimates closely matched the chromosome regions with a few exceptions. The upstream regions in *ebony* had estimates (π = .009, θ_w = .009) that were higher than those for the 3R Chromosome arm (π = .005, θ_w = .0063). The *Ddc* gene had lower estimates (π = .004, θ_w = .004) than for the 2L region (π = .006, θ_w = .008) in which it lies. Divergence of each candidate CDS sequence was also calculated in comparison to the consensus *D. simulans* gene (Table 2.13). Divergence estimates were highest for *pale* (.045) and the lowest for *Ddc* (.016). A sliding window analysis of π and θ_w was also used to measure these estimate across smaller segments along the genes (see above) and the upstream regions containing significant SNP sites (Table 2.14, Figure 2.6). Variation in nucleotide diversity across the genes and upstream regions was abundant. Some noteable peaks were seen in the upstream region of *ebony*, consistent with the higher π and θ_w estimates in this region compared to chromosome arm averages.

Measures of linkage disequilibrium (LD) (*D'* and R^2) among significant SNP sites revealed strong LD among significant SNP sites in different parts of the

candidate genes with the fewest sites in LD within *tan* (Figure 2.7, Figure 2.8). Sites up to 10 kb apart in *ebony* and its upstream region had significant estimates of LD. There was also clusters of nearby sites in *ebony* with strong LD, with some clusters appearing to segregate independently from one another. Some strong LD was also found between SNPs in *Ddc* up to about 4 kb apart and there was also some LD among sites located within shorter regions. *pale* and *tan* did not exhibit as strong LD across large distances and generally showed exponential rates of decay in D' and R^2 estimates with distance, with a few exceptions in *tan* (Figure 2.7, Figure 2.8).

Neutrality Tests

Tajima's D was calculated for each gene (exons and introns) as well as the 5kb upstream separately. While the estimates for each gene were negative, none of them were significantly different from zero (Table 2.12), indicating that the allele frequency spectrum of each gene may be as expected under drift-mutation equilibrium. The sliding window analysis did reveal several regions within and upstream of each gene with significant or marginally significant values but none contained sites associated with pigmentation (Table 2.14, Figure 2.6). Neutrality tests on the CDS regions of the candidate genes were conducted using coalescent simulations with *D. simulans* as the outgroup (Table 2.13). There were no significant values obtained for Tajimas D or the E-test. The genome wide average of D for the DGRP lines was -0.686 (Mackay, Richards et al. 2012) which is higher than most of the estimates of D for the candidate genes. Fay and Wu's H yielded significant results for Ddc ($H = -1.9$, $P < .05$)

and *tan* (H = -2.6, P <.05), which reflects an excess of high frequency derived SNPs that may be due to a selective sweep

A separate MK test was conducted on the coding regions for each gene using sequence data from the DGRP lines and the *D. simulans* sequence (Table 2.15). Departures from neutrality were found for *ebony* (G = 11.25, p <.001), but not for *Ddc* (G = 1.95,), *pale* (G = 1.11) or *tan*. Within *ebony* an excess of amino acid polymorphism was uncovered ($Pn/Ps > Dn/Ds$), with a neutrality index (*NI*) of 5.909, reflecting an excess of amino acid polymorphism which may be due to negative or purifying selection . Negative values for α were found for *ebony* as well (α= -4.9) which indicates possible sampling error or segregation of slightly deleterious amino acid substitutions.

The HKA test was performed to examine whether polymorphism levels and divergence across loci were correlated, as expected under neutral evolution. Coalescent simulations revealed significant departures of tested loci from neutrality (X^2 = 10.49, P < .05). This appears to be driven by significantly higher numbers of polymorphic sites and lower divergence rates in the coding regions between *D. melanogaster* and *D. simulans* for *pale* than under neutral expectations (Table 2.16). The results for *tan*, *Ddc* and *ebony* were not significantly different from values expected under the neutral model.

SNP sites associated with pigmentation in ebony

Among all the candidate genes, *ebony* had the most significant SNP sites associated with pigmentation traits genes. 31 out of the 212 tested SNPs in *ebony*

were significantly associated with pigmentation variation. Overall, the negative effect sizes for each SNP were larger in magnitude than the positive effect sizes, meaning on average they have a greater negative effect on luminosity and thus act to increase pigmentation intensity (Table 2.8, Table 2.9). However, there were a few sites that had a greater positive effect. Five of these sites were in the exons, while the rest were in introns or upstream of the start codon (Table 2.6). None of the coding changes resulted in amino acid polymorphisms. There was generally no overlap of significant SNPs between abdominal and thoracic traits except for a single site 422 bp upstream of the start codon (Table 2.7, Figure 2.5). *ebony* also had two SNP sites that were significantly associated with both expression and pigmentation (Table 2.10). One of these SNPs was located in in the seventh exon and the other was in the first intron.

Most of the SNPs in *ebony* that were associated with abdominal traits were located in regions upstream of the start codon. These sites are in a small ~630 bp region ~3 kb upstream and some demonstrated high and significant levels of LD with each other ($.71 < D' < 1$, $0.45 < R^2 < 1$, $P < .05$, Figure 2.7,Figure 2.8). Many of these sites had large negative effect sizes on abdominal traits (-.946, -1.093). For the thoracic traits, there was a wider spread of significant SNPs in the upstream region of *ebony* and in the first intron. The two main clusters in the first large intron (3.8 kb long) are about 1.8 kb apart and each cluster is less than 650 bp long. Within the group closest to the second exon (further downstream), there was significant LD among sites in this group, unlike the sites in the other cluster of sites in the intron.

Additionally, as stated above, there were significant SNP sites over 9kb apart in the *ebony* region that had high estimates for LD. There were also many sites shorter distances apart that were not in LD (Figure 2.7).

As mentioned earlier, nucleotide diversity estimates were noticeably higher for the 5k region upstream of *ebony* than for the 3R chromosome in which this gene resides. This upstream region contains 19 SNPs that were associated with pigmentation (Table 2.7, Figure 2.5). Sliding window analysis revealed that most of the sites upstream of the start codon were in regions with higher nucleotide diversity ($.009 < \pi < .017$, $.006 < \theta_w < .017$, Figure 2.6) than the estimates for the 3R Chromosome arm ($\pi = .0051$, $\theta_w = .0063$, Table 2.11). Additionally, one of the 650 bp regions in the first intron containing SNPs significantly associated with thoracic traits was also associated with high levels of diversity ($.008 < \pi < .011$, $.008 < \theta_w < .010$). Two of these sites had the highest negative values for effect sizes (-.773) on thoracic pigmentation. Overall, the estimates for the 500 kb windows in the sliding analysis of Tajima's D for *ebony* were negative (Table 2.14).

SNP sites associated with pigmentation in pale

In the *pale* gene region, there were ten SNPs out of the 141 tested sites that were significantly associated with pigmentation traits (Table 2.7). Overall, the positive effect sizes for each significant SNP were larger in magnitude than the negative effect sizes, meaning they tend to increase luminosity and therefore decrease pigmentation intensity (Table 2.8, Table 2.9). This trend was seen in the effect sizes

of every SNP site. Three sites were associated with abdominal pigmentation traits: two of these SNPs were located 53 bp apart in the 4th intron, while one was located upstream of the start codon. The two intronic SNPs are located in regions within a local peak with a high level of diversity (π = .014 and θ_w = .014). Site 1497 in the 4th intron had the largest effect size on abdominal pigmentation (.593).

The seven SNPs associated with thoracic traits were located within a 1 kb sequence located about 3056 bp upstream of the start codon of *pale* (Figure 2.5). SNP sites in this region had significant LD each other ($D' = 1$, $0.20 < R^2 < 0.90$, $P < 0.05$, Figure 2.7, Figure 2.8) and many were located within local peaks of high diversity ($.009 < \pi < .014$, $.006 < \theta_w < .011$, Table 2.14, Figure 2.6). However, these sites were not in LD with the site 305 bp downstream associated with abdominal pigmentation. Site 8721, which was the furthest site from the start codon, had the largest positive effect on thoracic pigmentation (0.620).

SNP sites associated with pigmentation in Ddc

The *Ddc* gene had 19 SNPs out of 121 sites that were associated with pigmentation. These were located in each of the two exons, in the single intron, and in the upstream region (Table 2.6). Overall, the negative effect sizes on luminosity for each SNP were larger in magnitude than the positive effect sizes, with an exception for abdominal cuticle pigmentation (Table 2.8, Table 2.9). The two polymorphisms in the second exon were associated with both sets of traits and did not result in amino acid changes. One of these sites (site 1061) had the highest positive effect sizes of alleles in this gene on both sets of abdominal traits (1.123, .787). The other

significant SNPs within the untranslated region of the first exon, spaced about 600 bp from each other, were associated with abdominal traits. Only a few of these sites showed significant LD between them. There were several SNPs located within 2.5 kb of the start codon associated with both sets of traits. Another cluster of significant SNPs was located about 4kb upstream of the start codon and associated with thoracic traits. Most of the SNPs in the upstream region, with the exception of two, showed significant LD with each other ($0.70 < D' < 1$, $0.12 < R^2 < 0.90$, $P < 0.05$) Significant values for LD were also found between significant SNP sites that were at least 4kb apart (Figure 2.7, Figure 2.8). Site 8148 in the upstream region had the highest negative effect size on thoracic pigmentation (-1.12). There were two SNP sites in *Ddc* that were associated with both thoracic and abdominal traits (Table 2.7, Table 2.14). One site in *Ddc* (Table 2.10) was associated with mRNA expression (Table 2.10). SNPs in the second exon ($\pi = .003$ and $\theta_w = .003$) and those more than 4kb upstream ($.0008 < \pi < .003$ and $.002 < \theta_w = .004$) were in areas exhibiting lower levels of diversity when compared to the estimates for the 2L chromosome arm ($\pi = .007$, $\theta_w = .008$).

SNP sites associated with pigmentation in tan

The *tan* gene had 11 out of 111 tested SNPs that were significantly associated with pigmentation, with the majority located in a region about 2.3 kb upstream of the start codon. Overall, the negative effect sizes on luminosity for each SNP were larger in magnitude than the positive effect sizes, with an exception for abdominal cuticle pigmentation (Table 2.8, Table 2.9). In the eighth exon, there was a polymorphism

that led to a change from alanine to threonine. The three significant SNPs within third intron were associated with abdominal traits and did not exhibit any significant LD with each other (Figure 2.8). All of the significant sites 1.1 kb upstream of the gene were associated with thoracic traits, with the majority being within a 1kb section that showed high LD ($0.56 < D' < 1$, $0.20 < R^2 < 1$, $P < 0.001$). Among these SNPS, Site 7939 and 8464 shared the highest negative effect sizes on both thoracic traits (-1.036, -1.340). There were two sites in tan that were significantly associated with expression; one in the 8th exon and one in the 3rd intron (Table 2.10). Sliding window analysis of nucleotide diversity revealed some variability within *tan* (Table 2.14, Figure 2.6).

Relationship between gene expression and pigmentation phenotypes

Expression data obtained from the DGRP was used determine whether adult female mRNA expression of pigmentation genes was associated with pigmentation phenotypic differences in the Raleigh isogenic lines (Table 2.17). Multivariate linear regression analysis indicated a significant negative relationship between adult expression of the *ebony* ($t = -2.833$, $P < 0.001$) and *pale* ($t = -2.363$, $P < 0.05$) genes and abdominal cuticle luminosity, with no effect contributed by *tan* or *Ddc*. This means that expression of each gene was positively correlated with pigmentation intensity. There also appeared to be a small positive interaction effect between the *pale* and *ebony* genes ($t = 2.242$, $P < 0.05$) on abdominal luminosity. This is indicative of a minor negative effect of dual expression in both *ebony* and *pale* on abdominal pigmentation.

The estimate of the coefficient or effect size was an order of magnitude higher for *pale* than *ebony* but the standard error (SE) was also an order of magnitude higher.

Discussion

Phenotypic variation within a natural population

The major aims of this study were to quantify population level variation of pigmentation traits in *Drosophila melanogaster* and determine how candidate genes in the pigmentation biosynthesis pathway may influence this variation. Phenotypic assays of the DGRP lines, which are largely isogenic and representative of a single population in Raleigh, NC, confirm the presence of significant within-population variation in both sets of measured thoracic and abdominal pigmentation traits. Although pigmentation traits within the two body parts were significantly correlated with each other, traits in different body parts had lower correlation coefficients, which suggest that there may be independent regulation of expression of pigmentation phenotypes. This is supported by estimates of genetic correlations among the traits, which were higher for traits on the same body part. This result also reflects that there are multiple loci simultaneously affecting different pigmentation traits but that these contributing loci also vary among the traits. Thus, these estimates of phenotypic and genetic correlations provide evidence for modular gene regulatory elements that affect different body parts, which is also supported by the SNP association results (see below).

The contribution of genetic variation to phenotypic variation

The SNP association study provided evidence that the pigmentation pathway genes *ebony*, *pale, tan*, and *Ddc* may be involved with variation in pigmentation traits in the DGRP lines from Raleigh, NC. There were SNPs in each gene that were

significantly associated with pigmentation variation. These genes are located on different chromosomes or chromosomal arms so it is highly unlikely any effects are due to linkage between the genes. However, it is possible that nearby linked genes may be influencing pigmentation variation instead. Linkage disequilibrium in *Drosophila* tends to decay past 1-2kb (Long, Lyman et al. 1998) so it is possible that some of the identified loci are indeed causative, since tested regions were much longer.

Previous work on *D. melanogaster* has already identified genes with influential effects on pigmentation phenotypes in different worldwide populations of *D. melanogaster*. For example, e*bony* has been implicated in sub-Saharan populations (Rebeiz, Pool et al. 2009) while *tan* and *bric-a-brac* have been implicated in European populations, with marginally significant results with *ebony* (Bastide, Betancourt et al. 2013). This suggests that geographic patterns on different continents can evolve independently with distinct genetic underpinnings. There may be sampling biases in these studies in the lines and genes tested, which may not exclude the potential for other genes to be acting in these populations. Otherwise, variation in identified genes may reflect the impact of the demographic history on these populations. *D. melanogaster* is thought to have undergone bottlenecks upon migration out of Africa, such that founder populations on different continents may possess distinct subsets of ancestral variation (David and Capy 1988, Baudry, Viginier et al. 2004). The results from this study are consistent with results found on African strains by Rebeiz & Poole (2009), which uncovered causative alleles in the upstream region of e*bony* associated

with abdominal pigmentation. Some of the SNP sites identified in this study were in a region that was identified as an abdominal and thoracic enhancer. Additional sites identified in this study may be due to de-novo mutations that arose since the migration of *D. melanogaster* out of Africa. Additionally, this is the first study, as far as we know, that has implicated *Ddc* and *pale* in pigmentation variation in *Drosophila melanogaster*.

Most of the significant SNP sites were in noncoding regions, which suggests that these maybe be located within transcriptional regulatory elements. This result is not particularly surprising, since these regions are expected to harbor more variation because they are not constrained by the need to maintain protein sequences. This was reflected in the estimates of nucleotide diversity (π, θ_w) being generally higher within in the 5kb regions upstream of each gene, with a single exception in *pale*. Previous work has identified potential regulatory regions within the genes included in this study that mediate expression of pigmentation and other traits. For example, regulatory elements have been identified in *ebony*, and causative SNPs sites have been uncovered in these elements that appear to regulate abdominal pigmentation (Rebeiz, Pool et al. 2009). Likewise, an regulatory element was identified within the intergenic region of two genes near *tan* that is required for regulation of abdominal and thoracic pigmentation in *D. yakuba* and *D. santomea* (Jeong, Rebeiz et al. 2008). Non-coding polymorphisms in *tan* have also been implicated in intra- and interspecific pigmentation variation between *D. novamexicana* and *D. americana*, (Jeong, Rebeiz et al. 2008, Wittkopp, Stewart et al.

2009, Bastide, Betancourt et al. 2013). Regulatory elements upstream of the promoter site in *Ddc* have been found to mediate spatiotemporal expression within the central nervous system (Hirsh, Morgan et al. 1986) and contribute to variation in longevity. These and linked sites may contribute to variation in levels of pigment biosynthesis.

Additionally, there was only a small overlap of SNP sites that were associated with both abdominal and thoracic traits inclusively, which suggests that different genetic components influence pigmentation expression in different tagmata. This reinforces results from the phenotypic assays that demonstrated possible independent regulation of pigmentation expression between the thorax and the abdomen. Regulatory work on pigmentation genes has already identified different enhancers that modulate expression in different body segments. For example, in addition to contributing to body pigmentation, there are enhancers in *yellow* that mediate production of wing spot formation in several species of *Drosophila* (Prud'homme, Gompel et al. 2006). Likewise, enhancers in *ebony* have been mapped to intronic and 5' regions that control gene expression in different body parts (Rebeiz, Pool et al. 2009). There is also evidence that the polymorphisms within and among the genes affect pigmentation traits differently, which is reflected in the variation in effect sizes among the alleles and genes.

SNP sites within pigmentation pathway genes

Ebony had the most significant SNP sites out of the tested candidate genes with significant associations to both abdominal and thoracic traits. It also had the highest estimates of nucleotide diversity for the gene and upstream regions. The

upstream SNP sites in *ebony* that are associated with abdominal pigmentation may correspond to regulatory elements that influence expression of pigmentation in the abdomen. These SNPs also appear to be in the same region (about 3.6 kb upstream) identified by Rebeiz, Pool et al. 2009 that was identified as an enhancer for abdominal pigmentation. There was another set of SNPs closer to the *ebony* promoter (2.8 kb upstream) that may be in an enhancer for thoracic pigmentation. These candidates SNPs were in regions with high estimates of nucleotide diversity that were substantially higher than estimates for the overall genome region, which suggests possible balancing selection acting on these upstream sites. Additionally there were two groupings of SNP sites in the first intron of *ebony* that were significantly associated with thoracic pigmentation. The large size (3.8 kb) of this intron makes it likely to harbor regulatory elements. The estimates of nucleotide diversity in the intron were not substantially different from the larger genomic region but were among the lowest within the *ebony* gene, which is unexpected for noncoding regions that typically are not constrained as coding regions. There was some strong LD between sites across the intron, indicating some haplotype structure. There was strong linkage across sites farther than 9 kb apart in the *ebony* region, which suggests some haplotype structure across *ebony* and its associated upstream region due to sites associated with pigmentation expression. However, there was also high nucleotide diversity across parts of the gene, which suggests the maintenance of older linked polymorphisms due to distant selective sweeps or the presence of inversions. The *ebony* gene may be within some of the documented inversion sites in *D. melanogaster*,

some which are present in the lines included in this study (Table 2.18, Table 2.19) (Huang, Massouras et al. 2014)

Most significant SNPs in *pale* were associated with thoracic traits and located in a 1 kb region about 3056 bp upstream of the transcription start site. These sites may be part of a regulatory element, and the high estimates of π and θ, and positive estimates for Tajima's D suggest that balancing selection may be acting in this 1kb region. Since the sites appear to be linked and have the same effect sizes it is not clear which of these sites may be causative. The SNPs in the fourth intron may be within another regulatory element which can be considered independent, since they are did not show significant LD with the upstream SNP sites. It is also possible that they are linked to unexamined sites downstream of *pale* that may be under selection. The *pale* gene had the most SNP sites that were associated with expression among all the tested candidate genes in this study. This suggests that these polymorphisms may be part of or linked to factors functionally involved with transcriptional regulation of the gene. However, none of these were also associated with pigmentation traits. This can be due to the central role of this gene and its dopamine products in multiple neurological functions such as behavior (Riemensperger, Isabel et al. 2011, Alekseyenko, Chan et al. 2013), locomotion (Pendleton, Rasheed et al. 2002), and circadian rhythmicity (Hirsh, Riemensperger et al. 2010). Additionally, expression was measured in flies 2-3 days post eclosion, a period in which *pale* expression is less likely to be pertinent to pigmentation patterning since this is set at a late pupal stage (Biessmann 1985).

Ddc had many SNPs significantly associated with pigmentation variation, with the majority being in potential regulatory regions upstream of the start codon. There was high LD among the SNP sites throughout the gene, especially in the upstream region, indicating some haplotype structure. It is possible that selection on some of these sites led to a sweep, a possibility that is supported by the valleys of nucleotide diversity in this region and significant negative Fay and Wu's H. The sweep may have not have been recent, since there are still high levels of polymorphism across these sites.

For the *tan* gene the majority of significant SNPs were in non-coding regions. The region 1.1 kb upstream from the promoter site may be an enhancer contributing to pigmentation expression in the thorax. This sequence is in the intergenic region of two nearby genes, which contained elements that may be required for regulation of abdominal and thoracic pigmentation in other *Drosophila* species (Jeong, Rebeiz et al. 2008). Similar to results in the other candidate genes, this putative enhancer was in a region of high nucleotide diversity, which may be due to balancing selection.

Gene expression and phenotypes

Differences in adult expression of *ebony* and *pale*, but not the other pigmentation genes, were associated with abdominal pigmentation differences, which suggest that transcription levels of these genes play a functional role in pigmentation variation. Expression was positively associated with luminosity, meaning that expression of these genes may be inversely correlated with pigmentation. *pale* encodes tyrosine hydroxylase, which is needed to convert tyrosine into pigmentation

precursors (Figure 2.1). *ebony* is not directly involved with melanin pigment production, but its step in the pathway affects melanization. It is thought that *ebony* expression is needed temporarily to store dopamine as NBAD before it is converted back to dopamine by *tan*. *tan* may not be a limiting factor in this process. *ebony* has been found to repress pigmentation expression (Wittkopp, True et al. 2002), which could explain the small interaction effect in the opposing direction between *ebony* and *pale* on abdominal pigmentation. No significant associations between these genes and thoracic pigmentation were found. These results may be confounded by the fact that expression data were collected using whole adult flies 2-3 days post-eclosion. Most pigmentation patterning is established in the late pupal stage (Biessmann 1985) while expression of *pale* and *ebony* have been shown to vary throughout ontogeny and in different body parts (Birman, Morgan et al. 1994, Hovemann, Ryseck et al. 1998, Pérez, Schachter et al. 2010)

Evolutionary Genetics of pigmentation pathway genes

This study allowed for an opportunity to examine the evolutionary dynamics within a set of genes that contribute to the same biosynthesis pathway. This pathway specifically involved in with expression of pigmentation phenotypes which is also the trait of interest. Theoretical and empirical work has demonstrated that the capacity for adaptive substitutions may vary among genes that are connected through a metabolic pathway (Flowers, Sezgin et al. 2007, Wright and Rausher 2010, Yu, Shen et al. 2011). Therefore an enzyme's placement in such a pathway may impact its evolutionary trajectory which can have downstream effects on quantitative traits

such as pigmentation. It is then important to understand these dynamics in order to understand how genetic variation may contribute to phenotypic diversity.

Pale is involved in the first step of the pigmentation biosynthesis pathway and is expected to be or have been under positive selection under the model proposed by Wright and Rausher (2010). However, from the present study it does not appear to be under positive selection, as reflected by nonsignificant results for the MK test. This is further supported by HKA results that demonstrated the coding region had lower levels of divergence and higher levels of diversity than expected under neutral evolution. The is due to higher level of polymorphisms in coding regions, which are mostly present in synonymous, or putatively neutral sites in the gene. Possible balancing selection may be acting on this gene and maintaining variation that originated previous to the split between *D. melanogaster* and *D. simulans*. This might be due to the role of this gene in other biological functions and pathways and possible complex selection on this gene. While there are some known and mapped inversions in this gene (Table 2.18), none were present in the DGRP lines included in this study (Table 2.19) (Huang, Massouras et al. 2014).

The MK test revealed an excess of nonsynonymous polymorphisms within the coding regions of *ebony* which may be due to negative or purifying selection and the segregation of slightly deleterious mutations. The negative values for Tajima's *D* support the possibility of purifying selection on these coding regions. This accumulation of slightly deleterious alleles may be due to less constraint on downstream genes. This pattern is expected if mutations affecting downstream

enzymes putatively have smaller effects (Wright and Rausher 2010). Additionally, previous work in *Drosophila* has suggested that a majority of non-synonymous mutations may have deleterious effects and may be subject to purifying selection (Loewe and Charlesworth 2006, Loewe, Charlesworth et al. 2006). The results for the other downstream genes, *Ddc* and *tan* were dissimilar and may reflect higher constraint since their activity is needed to produce Dopamine, which feeds into two branches of the pathway that produce two separate sets of pigment molecules. The significant negative values for Fay and Wu's *H* on the coding regions of *Ddc* and *tan* reflect possible selective sweep at these loci. These results are consistent with the evidence of positive selection on pathway branch points in *Drosophila* metabolic pathways (Flowers, Sezgin et al. 2007)

One source of variation that should be considered, but was not included in this present study are insertion mutations caused by transposable elements (TEs). Study of the DGRP lines revealed that these lines contained 36,810, TE with an average of 1,342 TEs per line, and 197,402 insertion sites total. There was data for 27 of the lines included in this study which average about 1542 TEs per line (Mackay, Richards et al. 2012). Another study used a high throughput approach with resequencing data from 166 strains of the DGRP lines to identify de novo insertion sites not present in the reference genome. Generally, the DGRP was found to contain about 8,000 new insertion sites, with de novo insertions for 38 TE families. (Linheiro and Bergman 2012). Another study, which included the DGRP lines, found that a majority of TE insertions were generally rare and found within single lines. Therefore, causative

TEs resulting in line specific effects may not contribute to population level variation in traits (Cridland, Macdonald et al. 2013). As technologies improve, the impact on TEs on phenotypic diversity can fully be characterized.

Conclusions

Overall, there is significant population level variation in thoracic and abdominal pigmentation traits in the DGRP lines, which are derived from a natural population in Raleigh, NC. Results from this work also provide evidence that various components of the pigmentation biosynthesis pathway have evolved together in this population and influence pigmentation expression. Most SNP sites that were significantly associated with pigmentation were found in non-coding regions and may be components of regulatory elements that effect expression of pigmentation genes. This also demonstrates the propensity for regulatory changes to generally contribute to variation in pigmentation phenotypes, which is supported by previous work on these genes and other morphological traits. I also characterized the level of segregating and divergent polymorphisms from *D. simulans* and estimated how selection may influence this variation. It is possible that the maintenance of genetic polymorphisms in pigmentation pathway genes may be an important factor in pigmentation trait variation in this population. This population is representative of individuals from Raleigh, NC, which is located in a temperate climate and is subject to seasonal changes in temperature and precipitation. Variation in selection related to seasonal fluctuations may favor the maintenance of pigmentation diversity and underlying genetic polymorphisms.

The results of this study have revealed the potential for multiple genetic components to contribute to phenotypic diversity. The identification of these candidate regions in pigmentation pathway genes has placed us in a position to identify adaptively important molecular variation in this species and to determine whether general genetic patterns underlie pigmentation evolution in *D. melanogaster* and possibly other species. Distinct associations of pigmentation gene variation with pigmentation in different parts of the adult fly underline the potential importance of modularity in gene expression in patterning different body parts independently. This study also provides an overview of how different components of a pathway may be evolving and how their positions within the biosynthesis pathway may influence selection on their genetic diversity.

References

Alekseyenko, O. V., Y.-B. Chan, R. Li and E. A. Kravitz (2013). "Single dopaminergic neurons that modulate aggression in *Drosophila*." Proceedings of the National Academy of Sciences **110**(15): 6151-6156.

Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollmann, L. H. Duncan, F. Lawrence, R. R. H. Anholt and T. F. C. Mackay (2009). "Systems genetics of complex traits in *Drosophila melanogaster*." Nat Genet **41**(3): 299-307.

Barnes, A. I. and M. T. Siva-Jothy (2000). "Density-dependent prophylaxis in the mealworm beetle *Tenebrio molitor* L. (Coleoptera: Tenebrionidae): cuticular melanization is an indicator of investment in immunity." Proc Biol Sci **267**(1439): 177-182.

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

Baudry, E., B. Viginier and M. Veuille (2004). "Non-African populations of *Drosophila melanogaster* have a unique origin." Molecular Biology and Evolution **21**(8): 1482-1491.

Benovoy, D., T. Kwan and J. Majewski (2008). "Effect of polymorphisms within probe–target sequences on olignonucleotide microarray experiments." Nucleic Acids Research **36**(13): 4417-4423.

Biessmann, H. (1985). "Molecular analysis of the *yellow* gene (*y*) region of *Drosophila melanogaster*." Proceedings of the National Academy of Sciences **82**(21): 7369-7373.

Birman, S., B. Morgan, M. Anzivino and J. Hirsh (1994). "A novel and major isoform of tyrosine hydroxylase in *Drosophila* is generated by alternative RNA processing." Journal of Biological Chemistry **269**(42): 26559-26567.

Chen, L., G. P. Page, T. Mehta, R. Feng and X. Cui (2009). "Single nucleotide polymorphisms affect both cis-and trans-eQTLs." Genomics **93**(6): 501-508.

Cochran, W. G. (1954). "Some methods for strengthening the common χ *2* tests." Biometrics **10**(4): 417-451.

Consortium, D. G. (2007). "Evolution of genes and genomes on the *Drosophila* phylogeny." Nature **450**(7167): 203-218.

Cridland, J. M., S. J. Macdonald, A. D. Long and K. R. Thornton (2013). "Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources." Molecular Biology and Evolution **30**(10): 2311-2327.

David, J. R. and P. Capy (1988). "Genetic variation of *Drosophila melanogaster* natural populations." Trends in Genetics **4**(4): 106-111.

David, J. R., P. Capy, P. Véronique and S. Tsakas (1985). "Thoracic trident pigmentation in *Drosophila melanogaster*: Differentiation of geographical populations." Genetics Selection Evolution **17**(2): 211-224.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.

Fay, J. C. and C.-I. Wu (2000). "Hitchhiking under positive Darwinian selection." Genetics **155**(3): 1405-1413.

Flowers, J., E. Sezgin, S. Kumagai, D. Duvernell, L. Matzkin, P. Schmidt and W. Eanes (2007). "Adaptive evolution of metabolic pathways in *Drosophila*." Molecular Biology and Evolution **24**(6): 1347-1354.

Futahashi, R., J. Sato, Y. Meng, S. Okamoto, T. Daimon, K. Yamamoto, Y. Suetsugu, J. Narukawa, H. Takahashi, Y. Banno, S. Katsuma, T. Shimada, K. Mita and H. Fujiwara (2008). "*yellow* and *ebony* are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*." Genetics **180**(4): 1995-2005.

Hegmann, J. P. and B. Possidente (1981). "Estimating genetic correlations from inbred strains." Behavior genetics **11**(2): 103-114.

Hey, J. (2001). HKA Program. Rutgers University.

Hirsh, J., B. A. Morgan and S. B. Scholnick (1986). "Delimiting regulatory sequences of the *Drosophila melanogaster Ddc* gene." Molecular and Cellular Biology **6**(12): 4548-4557.

Hirsh, J., T. Riemensperger, H. Coulom, M. Iché, J. Coupar and S. Birman (2010). "Roles of dopamine in circadian rhythmicity and extreme light sensitivity of circadian entrainment." Current Biology **20**(3): 209-214.

Hoekstra, H. E., J. G. Krenz and M. W. Nachman (2004). "Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations." Heredity **94**(2): 217-228.

Hovemann, B. T., R. P. Ryseck, U. Walldorf, K. F. Stortkuhl, I. D. Dietzel and E. Dessen (1998). "The *Drosophila ebony* gene is closely related to microbial peptide synthetases and shows specific cuticle and nervous system expression." Gene **221**(1): 1-9.

Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Rámia, A. Tarone, L. Turlapati, T. Zichner, D. Zhu, R. Lyman, M. Magwire, K. Blankenburg, M. A. Carbone, K. Chang, L. Ellis, S. Fernandez, Y. Han, G. Highnam, C. Hjelmen, J. Jack, M. Javaid, J. Jayaseelan, D. Kalra, S. Lee, L. Lewis, M. Munidasa, F. Ongeri, S. Patel, L. Perales, A. Perez, L. Pu, S. Rollmann, R. Ruth, N. Saada, C. Warner, A. Williams, Y.-Q. Wu, A. Yamamoto, Y. Zhang, Y. Zhu, R. Anholt, J. Korbel, D. Mittelman, D. Muzny, R. Gibbs, A. Barbadilla, S. Johnston, E. Stone, S. Richards, B. Deplancke and T. Mackay (2014). "Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines." Genome Research.

Hudson, R. R., M. Kreitman and M. Aguadé (1987). "A test of neutral molecular evolution based on nucleotide data." Genetics **116**(1): 153-159.

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

Joron, M., C. D. Jiggins, A. Papanicolaou and W. O. McMillan (2006). "*Heliconius* wing patterns: an evo-devo model for understanding phenotypic diversity." Heredity **97**(3): 157-167.

Kopp, A., I. Duncan and S. B. Carroll (2000). "Genetic control and evolution of sexually dimorphic characters in *Drosophila*." Nature **408**(6812): 553-559.

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

Kopp, A. and J. R. True (2002). "Evolution of male sexual characters in the Oriental *Drosophila melanogaster* species group." Evolution & Development **4**(4): 278-291.

Lavington, E., R. Cogni, C. Kuczynski, S. Koury, E. L. Behrman, K. R. O'Brien, P. S. Schmidt and W. F. Eanes (2014). "A Small System—high-resolution study of metabolic adaptation in the central metabolic pathway to temperate climates in *Drosophila melanogaster*." Molecular Biology and Evolution.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics **25**(11): 1451-1452.

Linheiro, R. S. and C. M. Bergman (2012). "Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*." PLoS One **7**(2): e30008.

Loewe, L. and B. Charlesworth (2006). "Inferring the distribution of mutational effects on fitness in *Drosophila*." Biology letters **2**(3): 426-430.

Loewe, L., B. Charlesworth, C. Bartolomé and V. Nöel (2006). "Estimating selection on nonsynonymous mutations." Genetics **172**(2): 1079-1092.

Long, A. D., R. F. Lyman, C. H. Langley and T. F. C. Mackay (1998). "Two sites in the delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*." Genetics **149**(2): 999-1017.

Lu, Y. and M. D. Rausher (2003). "Evolutionary Rate Variation in Anthocyanin Pathway Genes." Molecular Biology and Evolution **20**(11): 1844-1853.

Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire and J. M. Cridland (2012). "The *Drosophila melanogaster* genetic reference panel." Nature **482**(7384): 173-178.

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barron, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L.-L. Pu, C. Qu, M. Ramia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y.-Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman and R. A. Gibbs (2012). "The *Drosophila melanogaster* Genetic Reference Panel." Nature **482**(7384): 173-178.

Majerus, M. E. N. (1998). Melanism : evolution in action. Oxford ; New York, Oxford University Press.

Marygold, S. J., P. C. Leyland, R. L. Seal, J. L. Goodman, J. Thurmond, V. B. Strelets and R. J. Wilson (2013). "FlyBase: improvements to the bibliography." Nucleic Acids Res **41**(Database issue): D751-757.

Massouras, A., S. M. Waszak, M. Albarca-Aguilera, K. Hens, W. Holcombe, J. F. Ayroles, E. T. Dermitzakis, E. A. Stone, J. D. Jensen, T. F. C. Mackay and B. Deplancke (2012). "Genomic variation and its inpact on gene expression in *Drosophila melanogaster*." PLoS Genet **8**(11): e1003055.

McDonald, J. H. and M. Kreitman (1991). "Adaptive protein evolution at the *Adh* locus in *Drosophila*." Nature **351**(6328): 652-654.

Neckameyer, W. S. and K. White (1993). "*Drosophila* Tyrosine Hydroxylase is encoded by the *pale* locus." Journal of Neurogenetics **8**(4): 189-199.

Olson-Manning, C. F., C.-R. Lee, M. D. Rausher and T. Mitchell-Olds (2013). "Evolution of Flux Control in the Glucosinolate Pathway in *Arabidopsis thaliana*." Molecular Biology and Evolution **30**(1): 14-23.

Parkash, R., B. Kalra and V. Sharma (2010). "Impact of body melanisation on contrasting levels of desiccation resistance in a circumtropical and a generalist *Drosophila* species." Evolutionary Ecology **24**(1): 207-225.

Pendleton, R., A. Rasheed, T. Sardina, T. Tully and R. Hillman (2002). "Effects of Tyrosine Hydroxylase Mutants on Locomotor Activity in *Drosophila*: A Study in Functional Genomics." Behavior Genetics **32**(2): 89-94.

Pérez, M. M., J. Schachter, J. Berni and L. A. Quesada-Allué (2010). "The enzyme NBAD-synthase plays diverse roles during the life cycle of *Drosophila melanogaster*." Journal of Insect Physiology **56**(1): 8-13.

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

R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Rebeiz, M., J. E. Pool, V. A. Kassner, C. F. Aquadro and S. B. Carroll (2009). "Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population." Science **326**(5960): 1663-1667.

Riemensperger, T., G. Isabel, H. Coulom, K. Neuser, L. Seugnet, K. Kume, M. Iché-Torres, M. Cassar, R. Strauss, T. Preat, J. Hirsh and S. Birman (2011). "Behavioral consequences of dopamine deficiency in the *Drosophila* central nervous system." Proceedings of the National Academy of Sciences **108**(2): 834-839.

Rosenblum, E. B., H. E. Hoekstra and M. W. Nachman (2004). "Adaptive reptile color variation and the evolution of the *Mc1R* gene." Evolution **58**(8): 1794-1808.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Meth **9**(7): 671-675.

Tajima, F. (1983). "Evolutionary relationship of DNA sequences in finite populations." Genetics **105**(2): 437-460.

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.

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

Takahashi, A. Y. A. and T. Takano-Shimizu (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.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Molecular Biology and Evolution.

True, J. R. (2003). "Insect melanism: the molecules matter." Trends in Ecology & Evolution **18**(12): 640-647.

Watt, W. B. (1968). "Adaptive significance of pigment polymorphisms in *Colias* butterflies. I. Variation of melanin pigment in relation to thermoregulation." Evolution **22**(3): 437-458.

Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." Theoretical Population Biology **7**(2): 256-276.

Wittkopp, P. J. and P. Beldade (2009). "Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy." Seminars in Cell & Developmental Biology **20**(1): 65-71.

Wittkopp, P. J., S. B. Carroll and A. Kopp (2003). "Evolution in black and white: genetic control of pigment patterns in *Drosophila*." Trends in Genetics **19**(9): 495-504.

Wittkopp, P. J., E. E. Stewart, L. L. Arnold, A. H. Neidert, B. K. Haerum, E. M. Thompson, S. Akhras, G. Smith-Winberry and L. Shefner (2009). "Intraspecific polymorphism to interspecific divergence: Genetics of pigmentation in *Drosophila*." Science **326**(5952): 540-544.

Wittkopp, P. J., J. R. True and S. B. Carroll (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., K. Vaccaro and S. B. Carroll (2002). "Evolution of *yellow* gene regulation and pigmentation in *Drosophila*." Current Biology **12**(18): 1547-1556.

Wittkopp, P. J., B. L. Williams, J. E. Selegue and S. B. Carroll (2003). "*Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes." Proceedings of the National Academy of Sciences of the United States of America **100**(4): 1808-1813.

Wright, K. M. and M. D. Rausher (2010). "The evolution of control and distribution of adaptive mutations in a metabolic pathway." Genetics **184**(2): 483-502.

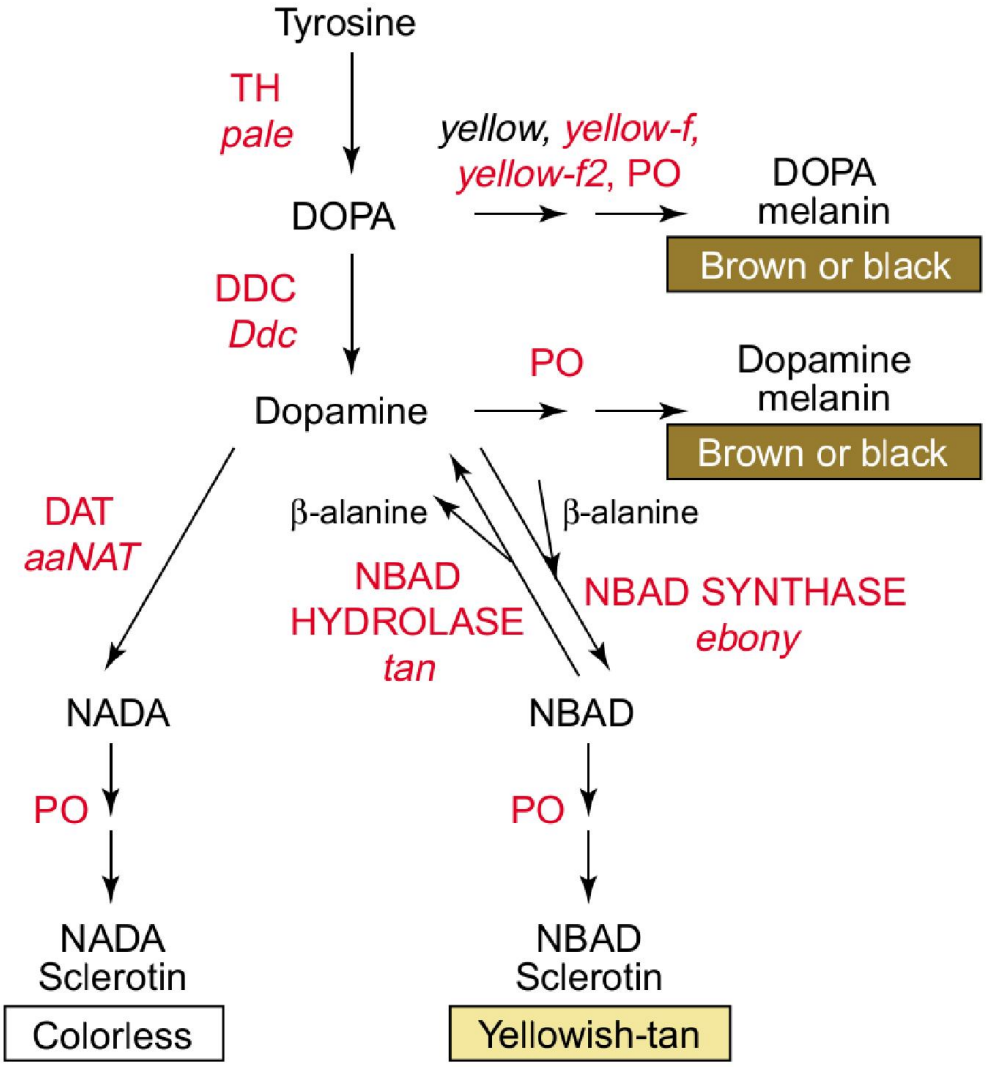
Wright, T. R. (1987). "The genetics of biogenic amine metabolism, sclerotization, and melanization in *Drosophila melanogaster*." Adv. Genet **24**(2): 127.

Yeh, S. D., S. R. Liou and J. R. True (2006). "Genetics of divergence in male wing pigmentation and courtship behavior between *Drosophila elegans* and *Drosophila gunungcola*." Heredity **96**(5): 383-395.

Yu, H.-S., Y.-H. Shen, G.-X. Yuan, Y.-G. Hu, H.-E. Xu, Z.-H. Xiang and Z. Zhang (2011). "Evidence of selection at melanin synthesis pathway loci during silkworm domestication." Molecular Biology and Evolution **28**(6): 1785-1799.

Zeng, K., Y.-X. Fu, S. Shi and C.-I. Wu (2006). "Statistical tests for detecting positive selection by utilizing high-frequency variants." Genetics **174**(3): 1431-1439.

Tables and Figures



s

Figure 2.1. Current model of the biosynthesis pathway and contributions of candidate genes for melanin pigment production. Red text indicates genes (italicized) and encoded enzymes (capitalized). Final pigments are at the ends of each pathway route with colored boxes representing pigment types. Abbreviations: DAT (dopamine acetyltransferase); DDC (DOPA decarboxylase); DOPA (dihydroxyphe-nylalanine); NADA (N-acetyl dopamine); NBAD(N-b-alanyl dopamine); PO, (phenoloxidases); TH(tyrosine hydroxylase).

Table 2.1. An overview of the candidate pigmentation genes investigated in this study and known roles in species divergence and intraspecific variation in studies of Drosophila

Gene (Protein)	Role in interspecific variation or species divergence in pigmentation traits	Proposed nature of genetic variation	References
ebony (NBAD Synthase)	Variation in abdominal pigmentation in African populations of *D. melanogaster*	regulatory	(Pool and Aquadro 2007), (Rebeiz, Pool et al. 2009)
	Variation in thoracic pigmentation in two wild derived strains of *D. melanogaster*	coding	(Takahashi, Takahashi et al. 2007)
	Clinal variation in thoracic trident pigmentation in east coast Australian populations of *D. melanogaster*	regulatory	(Telonis-Scott, Hoffmann et al. 2011)
	Interspecific variation in overall body color United States populations of *D. Americana*	regulatory	(Wittkopp, Stewart et al. 2009)
	Divergence in overall body color between *D. novamexicana* and *D. americana*	regulatory	(Wittkopp, Williams et al. 2003), (Wittkopp, Stewart et al. 2009)
pale (Tyrosine Hydroxylase)	Possible contribution to *Drosophila* wing melanin patterning differences		(True, Edwards et al. 1999)
yellow	Divergence in body pigmentation in *D. melanogaster*, *D. subobscura* , and *D. virilis*	regulatory	(Wittkopp, Vaccaro et al. 2002)
tan (NBAD Hydrolase)	Variation in abdominal pigmentation in European populations of *D. melanogaster*	regulatory	(Bastide, Betancourt et al. 2013)
	Variation in overall body color in United States populations of *D. Americana*	regulatory	(Wittkopp, Stewart et al. 2009)
	Divergence in overall body color between *D. novamexicana* and *D. Americana*	regulatory	(Wittkopp, Stewart et al. 2009)
	Abdominal pigmentation differences between *D. yakuba* and *D. santomea*	regulatory	(Jeong, Rebeiz et al. 2008)

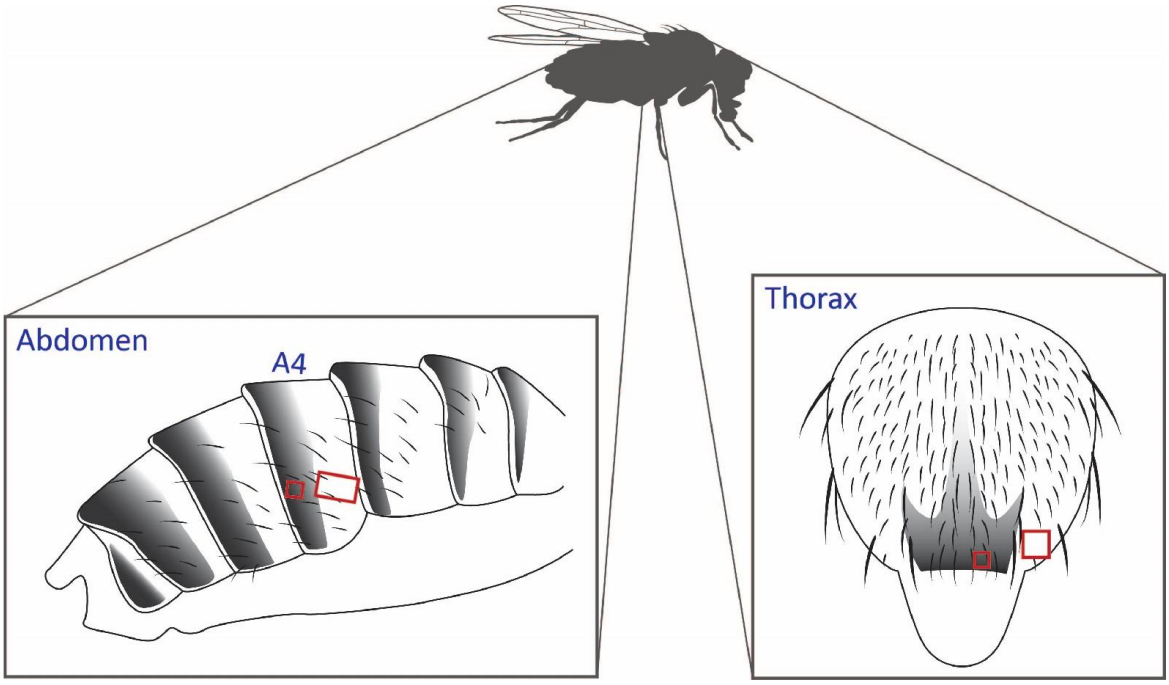


Figure 2.2. Diagram illustrating segments within thorax assayed for luminosity measurements. Red boxes indicates landmarks where measurements were taken in the A4 abdominal segment

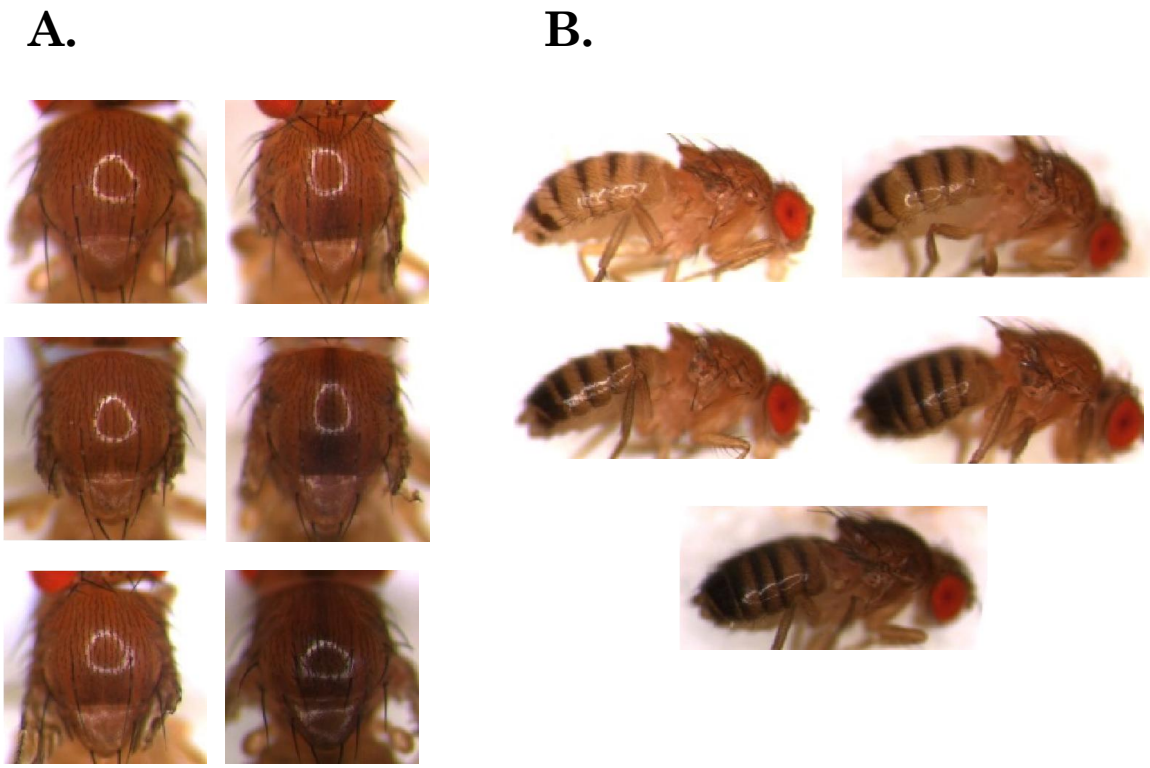


Figure 2.3. A. Representation of variation in female pigmentation from DGRP lines representing populations derived from Raleigh, North Carolina in both A. thoracic traits and B. abdominal traits.

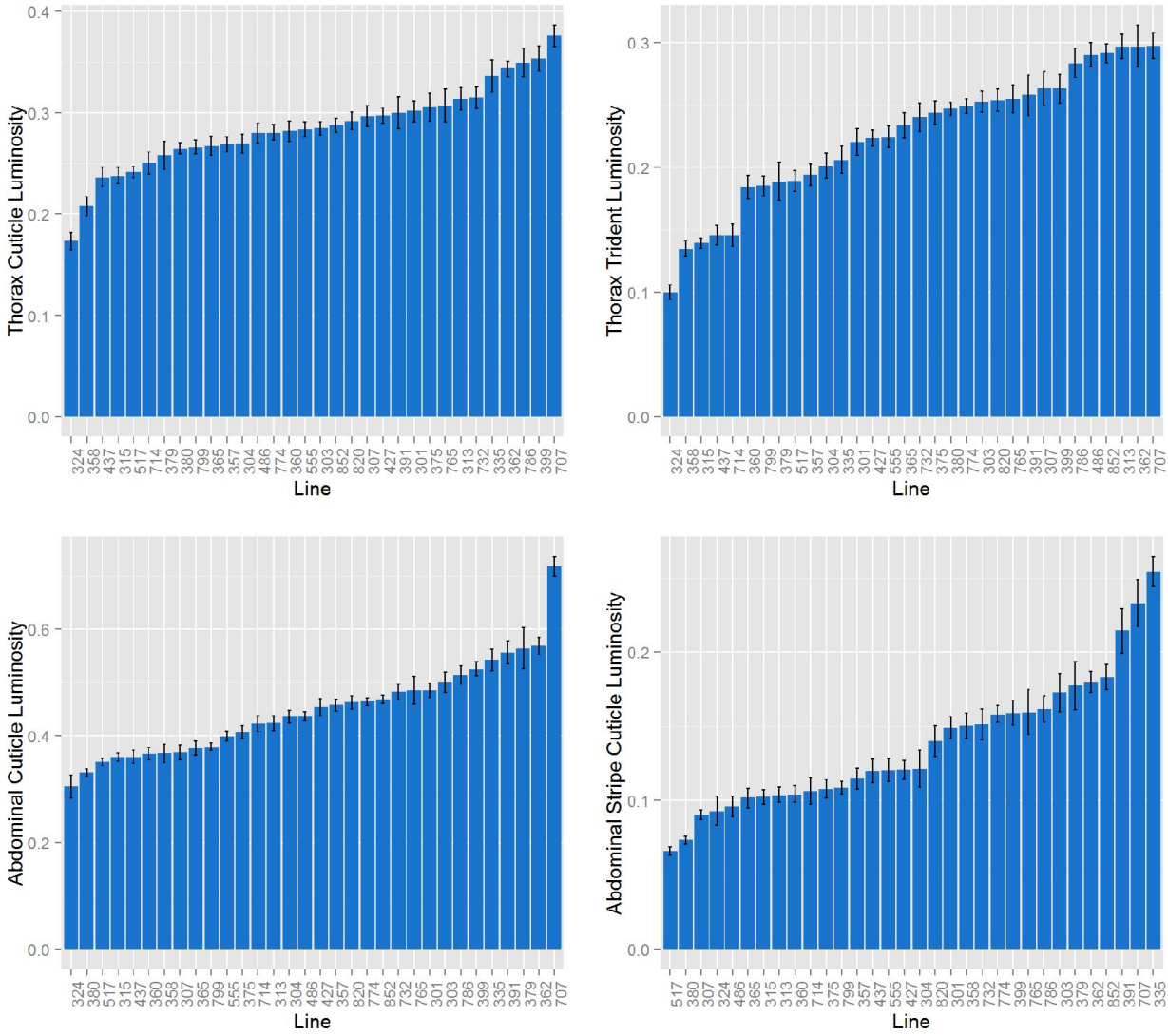


Figure 2.4. Box plots displaying mean luminosity scores and standard error bars for each DGRP line for all sets of measured pigmentation traits (Thoracic Cuticle, Thoracic Trident, Abdominal Cuticle, and Abdominal Stripe)

Table 2.2. Nested ANOVA Analysis of female pigmentation traits within DGRP lines representative of a single population from Raleigh, NC. Vials were nested within lines. . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Trait		**Df**	**SS**	**MS**	***F***	***P* - value**
Thorax Cuticle	Line	31	0.694	0.022388	7.700	**<.0001 *** **
	Vial	64	0.1861	0.002908		
	Error	288	0.3863	0.001341		
Thorax Trident	Line	31	1.0403	0.03356	15.460	**<.0001 *** **
	Vial	64	0.1389	0.00217		
	Error	288	0.3909	0.001357		
Abdominal Cuticle	Line	31	2.831	0.09133	8.582	**<.0001 *** **
	Vial	64	0.681	0.01064		
	Error	288	0.6638	0.002305		
Abdominal Stripe	Line	31	0.7145	0.023049	8.508	**<.0001 *** **
	Vial	64	0.1734	0.002709		
	Error	288	0.2645	0.000918		

Table 2.3. Phenotypic Correlation matrix for female pigmentation traits in Isogenic DGRP Lines from Raleigh, NC (N = 32). Bold values denote $P < 0.0001$. Significance was calculated with the Pearson method.

	Thorax Cuticle	Thorax Trident	Abdominal Cuticle	Abdominal Stripe
Thorax Cuticle				
Thorax Trident	**0.804*** **			
Abdominal Cuticle	**0.758*** **	**0.558*** **		
Abdominal Stripe	**0.340*** **	**0.340*** **	**0.851*** **	

Table 2.4. Estimation of genetic correlations of trait pairs as described in (Hegmann and Possidente 1981).

	Thorax Cuticle	Thorax Trident	Abdominal Cuticle	Abdominal Stripe
Thorax Cuticle				
Thorax Trident	0.3978			
Abdominal Cuticle	0.3753	0.2774		
Abdominal Stripe	0.2780	0.1636	0.4195	

Table 2.5a. Polymorphic sites within coding regions of *ebony* in *D. melanogaster* (n = 37) using *D. simulans* sequence as the reference. Site numbers refers to where the SNP sites resides within coding region of the gene. Highlighted sites denotes polymorphisms that had led to amino acid changes.

Site	57	75	99	136	171	189	210	232	234	246	264	297	298	343	372	378	391	402	454	459	463	465	474	480	504	506	582	624	651	663	700	702	705	708	711	715	720	723	729	735	741	750	754	756	770	789	805	811	852	855	858	864	865	894	903	924	954
D. Simulans	G	C	C	C	T	G	C	C	C	G	C	C	C	C	C	A	T	A	T	A	C	T	C	T	T	G	C	C	C	T	T	A	G	C	G	T	C	C	C	T	G	C	G	C	C	T	A	C	T	C	C	G	C	T	C	T	T
RAL-301	T	.	T	.	C	T	.	.	G	A	.	A	.	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	A	.	.	.	C	G	T	C	T	G	.	T	C	T	C	A
RAL-303	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-304	T	.	T	.	C	T	.	A	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-306	T	.	T	.	C	T	.	A	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-307	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	A	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-313	T	.	T	A	C	T	.	.	G	A	.	A	T	T	T	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	T	C	T	T	.	T	G	A	C	A	A	A	.	.	C	G	T	C	T	G	.	T	C	T	C	A
RAL-315	T	T	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	A	.	.	.	C	G	T	C	T	G	.	T	C	T	C	A
RAL-324	T	.	T	A	C	T	.	.	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-335	T	.	T	.	C	T	.	.	G	A	.	A	.	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-357	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	C	T	C	T	C	A
RAL-358	T	.	T	A	C	T	.	.	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-360	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	A	G	C	.	T	T	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-362	T	.	T	A	C	T	.	.	G	A	.	A	T	T	T	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	T	C	.	T	.	T	G	A	C	A	A	A	.	.	C	G	T	C	T	G	.	T	C	T	C	A
RAL-365	T	.	T	.	C	T	.	.	G	A	.	A	.	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-375	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	A	G	C	.	T	T	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	A	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-379	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	C	G	T	C	T	G	.	T	C	T	C	A
RAL-380	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-391	T	.	T	.	C	T	T	.	A	A	.	A	.	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-399	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	A	G	C	.	T	T	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	A	.	G	T	C	T	G	.	T	C	T	C	A
RAL-427	T	.	T	.	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	A	.	.	.	C	G	T	C	T	G	.	T	C	T	C	A
RAL-437	T	.	T	A	C	T	.	.	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-486	T	.	T	A	C	T	.	A	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-514	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-517	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	C	T	C	T	C	A
RAL-555	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-639	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	A	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-705	T	.	T	A	C	T	.	A	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-707	T	.	T	A	C	T	.	.	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-714	T	.	T	A	C	T	.	.	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-730	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-732	T	.	T	.	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-765	T	.	T	.	C	T	.	.	G	A	.	A	.	T	.	T	C	G	A	C	T	A	.	G	C	.	T	T	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	A	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-774	T	.	T	.	C	T	.	.	G	A	T	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	T	G	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-786	T	.	T	.	C	T	.	.	G	A	.	A	.	.	.	T	C	G	A	C	A	A	.	G	C	.	T	T	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	A	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-799	T	.	T	.	C	T	.	A	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	.	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A
RAL-820	T	.	T	A	C	T	.	.	G	A	.	A	T	T	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	C	T	C	T	C	A
RAL-852	T	.	T	.	C	T	T	.	G	A	.	A	T	.	.	T	C	G	A	C	T	A	.	G	C	.	T	.	G	C	C	.	C	.	.	.	T	G	A	C	A	.	.	.	.	G	T	C	T	G	.	T	C	T	C	A

Polymorphic sites in *ebony* coding regions(cont.)

Site	987	993	1044	1062	1116	1117	1141	1161	1182	1224	1233	1261	1263	1275	1287	1293	1299	1308	1326	1332	1347	1356	1365	1383	1386	1387	1395	1419	1423	1438	1467	1475	1479	1530	1539	1545	1551	1554	1560	1563	1572	1596	1611	1635	1644	1645	1662	1683	1686	1689	1737	1764	1788	1803	1815	1819
D. Simulans	C	C	T	C	T	C	C	A	T	C	T	C	G	G	G	C	A	C	C	C	A	C	A	G	T	C	G	A	C	A	C	C	C	A	C	T	T	A	C	C	A	T	G	A	G	T	C	C	T	C	T	C	C	C	C	T
RAL-301	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	T	G	C	C	T	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-303	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-304	T	.	C	G	A	T	T	G	G	T	C	T	A	T	T	T	T	T	.	A	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-306	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-307	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-313	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	.	G	C	C	.	A	G	T	C	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-315	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	T	G	C	C	T	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-324	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-335	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-357	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-358	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-360	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	.	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-362	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	T	.	C	.	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-365	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	T	G	C	C	T	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-375	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	T	.	C	.	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-379	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-380	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-391	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	.	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	A	A	.	T	C
RAL-399	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	T	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-427	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	T	G	C	C	T	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-437	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-486	T	.	C	G	A	T	T	G	G	T	C	T	A	T	T	T	T	T	.	A	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-514	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-517	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-555	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-639	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-705	T	T	C	G	A	T	T	G	G	T	C	T	A	T	T	T	T	T	.	A	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-707	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-714	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-730	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-732	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-765	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	.	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	T	T	G	C	C	G	A	A	T	T	.	.	A	.	A	T	T	C
RAL-774	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	T	.	C	.	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-786	T	T	C	G	A	.	T	G	G	T	C	.	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	T	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-799	T	T	C	G	A	T	T	G	G	T	C	T	A	.	T	T	T	T	.	A	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C
RAL-820	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	.	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	C	T	A	.	A	.	T	C
RAL-852	T	T	C	G	A	.	T	G	G	T	C	T	A	.	T	T	T	T	.	A	C	T	G	C	C	.	A	G	.	.	G	.	T	G	T	C	C	G	.	T	G	C	C	G	A	A	T	T	.	.	A	.	A	.	T	C

Polymorphic sites in *ebony* coding regions (cont.)

Site	1821	1833	1834	1839	1848	1866	1896	1911	1932	1949	1953	1965	1992	1996	2010	2038	2040	2052	2085	2103	2124	2133	2169	2220	2235	2238	2253	2255	2262	2271	2289	2337	2364	2383	2418	2427	2433	2451	2523	2532	2547	2556	2565	2610	2619	2633	2637
D. Simulans	**G**	**A**	**C**	**C**	**T**	**T**	**G**	**A**	**G**	**A**	**C**	**G**	**G**	**T**	**G**	**G**	**G**	**C**	**G**	**C**	**T**	**T**	**G**	**G**	**A**	**G**	**G**	**A**	**A**	**G**	**C**	**G**	**G**	**C**	**C**	**T**	**G**	**G**	**C**	**C**	**G**	**C**	**A**	**C**	**C**	**C**	**G**
RAL-301	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	.	G	C	A	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-303	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-304	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	A	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-306	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-307	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-313	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	C	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	.	T	C	T	A	T	.	.	.	C	T	T	G	A
RAL-315	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	.	G	C	A	.	G	A	T	T	.	.	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-324	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-335	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-357	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	A	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-358	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-360	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	A	T	.	.	.	C	T	T	G	A
RAL-362	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	C	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	.	.	.	C	T	T	G	A
RAL-365	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	.	G	C	A	.	G	A	T	T	.	.	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-375	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	.	T	C	T	A	T	.	.	.	C	T	T	G	A
RAL-379	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-380	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	.	.	.	C	T	T	G	A
RAL-391	.	G	T	T	C	C	A	G	A	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	T	G	A	T	T	.	.	T	C	T	.	T	.	.	.	C	T	T	G	A
RAL-399	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	.	.	.	C	T	T	G	A
RAL-427	.	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	.	G	C	A	.	G	A	T	T	.	.	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-437	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-486	A	G	T	T	C	C	A	G	.	G	T	A	.	C	.	.	A	T	A	.	C	A	C	A	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	T	.	C	T	T	G	A
RAL-514	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-517	.	G	T	T	C	C	A	G	.	.	T	A	.	C	T	.	A	T	.	.	C	A	C	.	G	C	A	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-555	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-639	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-705	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-707	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-714	A	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-730	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-732	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	T	T	.	C	T	T	G	A
RAL-765	A	G	T	T	C	C	A	G	.	.	T	A	A	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	.	T	C	T	.	T	.	.	T	C	T	T	G	A
RAL-774	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	T	T	C	T	.	T	.	.	.	C	T	T	G	A
RAL-786	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	T	C	A	C	A	G	C	.	.	G	A	T	T	.	T	T	C	T	A	T	.	.	.	C	T	T	G	A
RAL-799	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	A	.	C	A	C	.	G	C	A	.	G	A	T	T	.	.	T	C	T	.	T	.	.	.	C	T	T	G	A
RAL-820	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	T	T	T	C	T	.	T	T	.	.	C	T	T	G	A
RAL-852	.	G	T	T	C	C	A	G	.	.	T	A	.	C	.	.	A	T	.	.	C	A	C	.	G	C	.	.	G	A	T	T	.	.	T	C	T	A	T	.	.	.	C	T	T	G	A

Table 2.5b. Polymorphic sites within coding regions of *pale* in *D. melanogaster* (n = 37) using *D. simulans* sequence as the reference. Site numbers refers to where the SNP sites resides within coding region of the gene. Highlighted sites denotes polymorphisms that had led to amino acid changes.

Site	165	189	190	195	206	228	346	424	453	512	540	576	586	588	598	603	630	642	657	663	665	708	712	714	753	759	783	831	846	852	855	858	867	906	912	942	993	996	1011	1014	1059	1098	1137	1188	1215	1284	1344	1383	1422	1432	1458	1482	1500	1593	1599	1638	1653	1668	1737
D. Simulans	**C**	**T**	**C**	**G**	**T**	**C**	**C**	**C**	**A**	**C**	**G**	**C**	**A**	**G**	**G**	**C**	**C**	**G**	**T**	**C**	**A**	**C**	**C**	**G**	**G**	**C**	**G**	**T**	**C**	**G**	**C**	**C**	**C**	**C**	**C**	**C**	**G**	**C**	**C**	**G**	**A**	**G**	**C**	**C**	**C**	**G**	**A**	**C**	**C**	**C**	**C**	**C**	**G**	**G**	**G**	**G**	**G**	**C**	**C**
RAL-301	.	C	.	.	.	.	G	T	.	.	.	.	G	.	.	T	T	A	.	.	G	T	.	.	.	.	A	.	T	.	T	.	.	.	.	.	.	.	.	.	G	.	.	T	A	.	.	T	.	.	T	A	C	.	A	.	C	T	.
RAL-303	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-304	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	.	.	.	G	T	.	.	.	.	A	.	A	T	.	A	.	T	.	A	T	T	.	G	.	.	T	.	A	T	T	.	T	.	.	C	.	A	.	C	T	.
RAL-306	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-307	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	.	T	.	.	C	.	A	.	C	T	.
RAL-313	.	C	.	A	.	.	G	T	.	.	.	T	G	.	.	.	.	.	C	T	G	T	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	.	G	.	.	T	.	.	.	T	.	.	T	A	C	.	A	.	.	T	T
RAL-315	.	C	.	A	.	.	G	T	.	.	A	.	G	.	.	T	.	.	.	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	T	.	T	.	C	.	A	.	C	T	.
RAL-324	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-335	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-357	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-358	.	C	.	A	.	.	G	T	.	.	.	.	G	.	.	.	.	.	C	T	G	T	.	.	.	.	.	.	.	.	T	T	A	.	.	.	T	.	.	G	.	.	T	.	.	.	T	.	.	T	A	C	.	A	.	.	T	T
RAL-360	.	C	.	A	C	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-362	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	.	.	G	T	.	.	.	A	.	A	T	.	A	.	T	.	A	T	T	.	G	.	T	T	.	A	.	T	.	.	T	A	C	.	A	.	.	T	.
RAL-365	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	A	.	G	.	.	.	.	G	C	T	T	T	.	T	T	.	T	.	.	C	.	A	.	C	T	.
RAL-375	.	C	.	A	.	.	G	T	.	.	A	.	G	.	.	T	T	.	.	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	T	.	T	.	C	.	A	.	C	T	.
RAL-379	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	A	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-380	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-391	.	C	.	A	.	.	G	T	.	.	.	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	.	T	.	A	C	.	A	.	C	T	.
RAL-399	.	C	.	A	.	.	G	T	.	.	.	.	G	.	.	T	T	A	.	.	G	T	T	.	.	.	.	.	.	.	A	T	.	.	.	T	G	.	T	.	A	G	C	T	T	A	.	.	T	.	.	T	.	C	.	A	A	.	T	.
RAL-427	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-437	.	C	.	A	.	.	G	T	.	.	.	.	G	.	A	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-486	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	.	T	.	A	C	.	A	.	C	T	.
RAL-514	.	C	.	A	.	.	G	T	.	.	.	T	G	.	.	.	.	.	C	.	G	T	T	.	.	.	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	T	.	C	C	A	.	C	T	.
RAL-517	.	C	.	A	.	.	G	T	.	.	.	T	G	.	.	.	.	.	C	.	G	T	T	.	.	.	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-555	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-639	.	C	.	A	.	.	G	T	.	.	.	T	G	.	.	.	.	.	C	.	G	T	T	.	.	.	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-705	.	C	.	.	.	.	G	T	.	.	.	.	G	.	.	T	T	A	.	.	G	T	.	.	.	A	.	T	.	T	.	.	.	.	.	.	.	.	.	.	G	.	.	T	A	.	.	T	.	.	T	A	C	.	A	.	C	T	.
RAL-707	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	T	.	T	.	C	.	A	.	C	T	.
RAL-714	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-730	.	C	.	.	.	T	G	T	.	.	.	.	G	T	.	.	.	.	.	.	G	T	.	A	.	.	.	A	.	.	T	T	.	.	.	.	.	T	T	.	G	.	.	T	.	.	.	T	.	.	T	A	C	.	A	.	C	T	.
RAL-732	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	.	T	.	A	C	.	A	.	C	T	.
RAL-765	T	C	T	A	.	.	G	T	.	.	A	.	G	.	.	T	T	.	.	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	T	A	C	.	A	.	.	T	.
RAL-774	.	C	.	A	.	.	G	T	.	.	.	T	G	.	.	.	.	.	C	T	G	T	.	.	.	.	.	.	.	.	T	T	.	.	.	.	T	.	.	G	.	.	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-786	.	C	.	A	.	.	G	T	T	G	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.
RAL-799	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	.	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	.	T	.
RAL-820	.	C	.	A	.	.	G	T	.	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	T	T	.	T	.	A	C	.	A	.	C	T	.
RAL-852	.	C	.	A	.	.	G	T	T	.	A	T	G	.	.	.	.	.	C	.	G	T	T	.	A	T	.	A	.	A	T	.	A	.	.	G	.	.	.	.	G	C	T	T	.	.	.	T	.	.	.	A	C	.	A	.	C	T	.

Table 2.5c. Polymorphic sites within coding regions of *Ddc* in *D. melanogaster* (n = 37) *using D. simulans* sequence as the reference. Site numbers refers to where the SNP sites resides within coding region of the gene. Highlighted sites denotes polymorphisms that lead led to amino acid changes.

Site	34	45	147	189	226	237	240	291	312	333	375	441	495	516	588	615	643	652	669	672	684	696	705	726	771	790	798	801	819	876	918	939	975	984	1002	1029	1062	1077	1173	1197	1206	1263	1290	1308	1311	1389
D. Simulans	**C**	**C**	**G**	**C**	**C**	**C**	**G**	**G**	**G**	**C**	**G**	**G**	**G**	**C**	**C**	**G**	**C**	**C**	**C**	**T**	**C**	**T**	**C**	**C**	**G**	**C**	**G**	**C**	**C**	**T**	**C**	**C**	**T**	**G**	**C**	**C**	**C**	**C**	**T**	**A**	**A**	**A**	**C**	**T**	**C**	**A**
RAL-301	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-303	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-304	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-306	A	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	.	G	A	T	G	A	T	T	C	A	T	T	T	G	A	G	.	G	T	C	T	C
RAL-307	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-313	.	T	C	T	.	T	T	A	A	T	T	A	A	.	.	.	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-315	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	.	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-324	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-335	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-357	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-358	A	T	C	T	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-360	.	T	C	.	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-362	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	.	G	A	T	G	A	T	T	C	A	T	T	T	G	A	G	.	G	T	C	T	C
RAL-365	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-375	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-379	A	T	C	.	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-380	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	.	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-391	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-399	A	T	C	.	T	T	T	A	A	T	T	.	A	.	A	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-427	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-437	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-486	A	T	C	.	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-514	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-517	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-555	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-639	A	T	C	.	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-705	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-707	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	.	A	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-714	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-730	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	A	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-732	A	T	C	T	T	T	T	A	A	T	T	.	A	.	.	.	T	T	G	C	A	G	T	T	.	A	A	T	G	A	T	.	C	A	T	T	T	G	A	G	.	G	T	C	T	C
RAL-765	.	T	C	T	.	T	T	A	A	T	T	.	A	G	.	C	T	T	G	C	A	G	T	T	.	G	A	T	G	A	T	T	C	A	T	T	T	G	A	G	.	G	T	C	T	C
RAL-774	.	T	C	T	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-786	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-799	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-820	A	T	C	.	.	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C
RAL-852	A	T	C	.	T	T	T	A	A	T	T	.	A	.	.	C	T	T	G	C	A	G	T	T	A	G	.	T	G	A	T	.	C	A	T	T	T	G	A	G	T	G	T	C	T	C

Table 2.5d. Polymorphic sites within coding regions of *tan* in *D. melanogaster* (n = 37) using *D. simulans* sequence as the reference. Site numbers refers to where the SNP sites resides within coding region of the gene. Highlighted sites denotes polymorphisms that lead led to amino acid changes.

	21	66	69	90	174	192	207	228	357	366	423	462	474	495	501	525	534	555	564	570	577	585	627	636	714	729	738	741	765	801	816	825	837	882	885	927	939	946	1011	1068	1095	1098	1107	1150
D. Simulans	G	C	C	C	C	C	A	G	C	T	G	G	C	C	C	C	C	C	G	C	C	C	A	C	A	C	C	T	G	G	C	C	G	C	G	G	C	T	C	G	G	C	C	C
RAL-301	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-303	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-304	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-306	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-307	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-313	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-315	.	T	.	T	A	.	T	T	A	C	A	A	.	T	T	T	.	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	T	T	A	C	A	G	T	T	C	T	T	T
RAL-324	.	T	.	T	G	T	T	C	A	C	A	A	T	.	G	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	A	C	A	G	T	T	C	T	T	T
RAL-335	.	T	.	T	G	T	T	C	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	A	C	A	G	T	T	C	T	T	.
RAL-357	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-358	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-360	.	T	.	T	G	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-362	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-365	.	T	.	T	G	T	T	C	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-375	A	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-379	.	.	T	T	.	T	T	C	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-380	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-391	.	T	.	T	A	T	T	C	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-399	.	T	.	T	G	T	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-427	.	T	.	T	G	T	T	C	A	C	A	A	T	T	T	T	.	T	T	.	T	T	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-437	.	T	.	T	A	.	T	T	A	C	A	A	T	T	T	T	.	T	T	.	T	T	.	A	T	T	T	C	A	C	T	T	.	T	A	C	A	G	T	T	C	T	T	T
RAL-486	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-514	A	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-517	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-555	.	T	.	T	G	T	T	C	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	.
RAL-639	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	A	C	A	G	T	T	C	T	T	T
RAL-705	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-707	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-714	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-730	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-732	.	T	.	T	G	T	T	C	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-765	.	T	.	T	A	.	T	T	A	C	A	A	T	T	T	T	.	T	T	.	T	T	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-774	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	A	C	A	G	T	T	C	T	T	T
RAL-786	.	T	.	T	A	.	T	T	A	C	A	A	T	.	G	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-799	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	.	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-820	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T
RAL-852	.	T	.	T	A	.	T	T	A	C	A	A	T	.	T	.	T	T	T	T	T	.	G	A	T	T	T	C	A	C	T	T	.	T	.	C	A	G	T	T	C	T	T	T

Table 2.6. Summary of candidate genes, tested SNP sites and significant SNP sites. Length includes gene region (exons and introns) as well as the 5lb upstream region.

				Significant SNP Sites			
Gene	**Chromosome Location**	**Length (bp)**	**Tested SNP Sites**	**Coding**	**Non-Coding**	**Upstream**	**Total**
ebony	3R	12339	212	5	10	19	34
pale	3L	10488	141	0	2	8	10
Ddc	2L	8824	121	6	2	11	19
tan	X	10603	116	1	3	7	11

Table 2.7. List of SNP sites that have significant associations with measured pigmentation traits following the permutation test. Position refers to where the SNP sites resides within the tested gene region starting at the end position of each gene and ending 5kb upstream of the gene. . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Gene	Position	Region	*F*	*P-value*	Trait
ebony	1671	exon 4	42.20747	<.0001***	Thorax Cuticle
ebony	1997	exon 3	33.77935	<.0001***	Thorax Cuticle
ebony	3586	intron 1	31.86957	<.0001***	Thorax Cuticle
ebony	3885	intron 1	31.86957	<.0001***	Thorax Cuticle
ebony	6245	intron 1	45.17918	<.0001***	Thorax Cuticle
ebony	10143	upstream	35.65527	<.0001***	Thorax Cuticle
ebony	10233	upstream	29.63892	<.0001***	Thorax Cuticle
ebony	10313	upstream	55.10596	<.0001***	Thorax Cuticle
ebony	10702	upstream	30.73713	<.0001***	Thorax Cuticle
ebony	11116	upstream	29.39542	<.0001***	Thorax Cuticle
ebony	11561	upstream	42.87062	<.0001***	Thorax Cuticle
ebony	312	exon 7	86.48501	<.0001***	Thorax Trident
ebony	1671	exon 4	71.08181	<.0001***	Thorax Trident
ebony	2480	exon 2	65.65824	<.0001***	Thorax Trident
ebony	3503	intron 1	57.17681	<.0001***	Thorax Trident
ebony	3586	intron 1	101.2314	<.0001***	Thorax Trident
ebony	3761	intron 1	95.93686	<.0001***	Thorax Trident
ebony	3885	intron 1	101.2314	<.0001***	Thorax Trident
ebony	4130	intron 1	71.85922	<.0001***	Thorax Trident
ebony	5134	intron 1	63.73248	<.0001***	Thorax Trident
ebony	5825	intron 1	70.72067	<.0001***	Thorax Trident
ebony	6182	intron 1	86.49624	<.0001***	Thorax Trident
ebony	6245	intron 1	139.9178	<.0001***	Thorax Trident
ebony	6450	intron 1	76.46055	<.0001***	Thorax Trident
ebony	10233	upstream	85.36719	<.0001***	Thorax Trident
ebony	10246	upstream	72.84752	<.0001***	Thorax Trident
ebony	10525	upstream	71.00491	<.0001***	Thorax Trident
ebony	10702	upstream	87.59508	<.0001***	Thorax Trident
ebony	11116	upstream	80.77845	<.0001***	Thorax Trident
ebony	9382	upstream	34.37399	<.0001***	Abdominal Cuticle
ebony	10935	upstream	47.8191	<.0001***	Abdominal Cuticle
ebony	11144	upstream	34.37399	<.0001***	Abdominal Cuticle
ebony	11397	upstream	34.37399	<.0001***	Abdominal Cuticle
ebony	11449	upstream	34.37399	<.0001***	Abdominal Cuticle
ebony	11473	upstream	34.37399	<.0001***	Abdominal Cuticle
ebony	11484	upstream	34.37399	<.0001***	Abdominal Cuticle

ebony	2882	exon 2	50.60872	<.0001***	Abdominal Stripe
ebony	9309	upstream	37.08905	<.0001***	Abdominal Stripe
ebony	9382	upstream	53.55908	<.0001***	Abdominal Stripe
ebony	10935	upstream	63.50495	<.0001***	Abdominal Stripe
ebony	10943	upstream	42.57823	<.0001***	Abdominal Stripe
ebony	11144	upstream	53.55908	<.0001***	Abdominal Stripe
ebony	11167	upstream	59.94156	<.0001***	Abdominal Stripe
ebony	11201	upstream	57.95464	<.0001***	Abdominal Stripe
ebony	11397	upstream	53.55908	<.0001***	Abdominal Stripe
ebony	11449	upstream	53.55908	<.0001***	Abdominal Stripe
ebony	11473	upstream	53.55908	<.0001***	Abdominal Stripe
ebony	11484	upstream	53.55908	<.0001***	Abdominal Stripe
ebony	11561	upstream	42.83917	<.0001***	Abdominal Stripe
pale	8568	upstream	33.34227	<.0001***	Thorax Cuticle
pale	8587	upstream	33.34227	<.0001***	Thorax Cuticle
pale	8615	upstream	33.34227	<.0001***	Thorax Cuticle
pale	8544	upstream	70.88261	<.0001***	Thorax Trident
pale	8568	upstream	87.6939	<.0001***	Thorax Trident
pale	8587	upstream	87.6939	<.0001***	Thorax Trident
pale	8615	upstream	87.6939	<.0001***	Thorax Trident
pale	8635	upstream	66.51538	<.0001***	Thorax Trident
pale	8873	upstream	84.89223	<.0001***	Thorax Trident
pale	9550	upstream	70.93438	<.0001***	Thorax Trident
pale	1444	intron 4	46.8747	<.0001***	Abdominal Cuticle
pale	1497	intron 4	34.92847	<.0001***	Abdominal Cuticle
pale	7187	upstream	36.5359	<.0001***	Abdominal Cuticle
pale	1444	intron 4	36.31037	<.0001***	Abdominal Stripe
Ddc	1034	exon 2	51.53486	<.0001***	Thorax Cuticle
Ddc	1061	exon 2	50.81493	<.0001***	Thorax Cuticle
Ddc	4338	upstream	39.8805	<.0001***	Thorax Cuticle
Ddc	4351	upstream	39.8805	<.0001***	Thorax Cuticle
Ddc	5529	upstream	46.41011	<.0001***	Thorax Cuticle
Ddc	8148	upstream	60.82772	<.0001***	Thorax Cuticle
Ddc	8449	upstream	36.74111	<.0001***	Thorax Cuticle
Ddc	8662	upstream	40.33372	<.0001***	Thorax Cuticle
Ddc	8721	upstream	59.11466	<.0001***	Thorax Cuticle
Ddc	2160	intron 1	64.37839	<.0001***	Thorax Trident
Ddc	2784	exon 1	66.47594	<.0001***	Thorax Trident
Ddc	5529	upstream	60.29552	<.0001***	Thorax Trident
Ddc	8148	upstream	72.32355	<.0001***	Thorax Trident
Ddc	1034	exon 2	37.69236	<.0001***	Abdominal Cuticle
Ddc	1061	exon 2	72.27385	<.0001***	Abdominal Cuticle

Ddc	2784	exon 1	38.16119	<.0001***	Abdominal Cuticle
Ddc	3063	exon 1	38.49731	<.0001***	Abdominal Cuticle
Ddc	5440	upstream	34.99909	<.0001***	Abdominal Cuticle
Ddc	6000	upstream	37.94844	<.0001***	Abdominal Cuticle
Ddc	1061	exon 2	33.89329	<.0001***	Abdominal Stripe
Ddc	1941	intron 1	32.3429	<.0001***	Abdominal Stripe
Ddc	2784	exon 1	61.91305	<.0001***	Abdominal Stripe
Ddc	2797	exon 1	56.46683	<.0001***	Abdominal Stripe
Ddc	3063	exon 1	36.00674	<.0001***	Abdominal Stripe
Ddc	3371	exon 1	37.17956	<.0001***	Abdominal Stripe
Ddc	3828	upstream	38.70079	<.0001***	Abdominal Stripe
Ddc	4957	upstream	32.50969	<.0001***	Abdominal Stripe
Ddc	6000	upstream	36.00939	<.0001***	Abdominal Stripe
tan	3408	intron 3	30.17248	<.0001***	Thorax Cuticle
tan	6716	upstream	32.62689	<.0001***	Thorax Cuticle
tan	7939	upstream	48.32766	<.0001***	Thorax Cuticle
tan	8464	upstream	47.45035	<.0001***	Thorax Cuticle
tan	860	exon 8	98.19661	<.0001***	Thorax Trident
tan	7939	upstream	132.487	<.0001***	Thorax Trident
tan	7945	upstream	64.43579	<.0001***	Thorax Trident
tan	8218	upstream	103.2173	<.0001***	Thorax Trident
tan	8464	upstream	122.4392	<.0001***	Thorax Trident
tan	8551	upstream	77.51238	<.0001***	Thorax Trident
tan	8996	upstream	78.11605	<.0001***	Thorax Trident
tan	2262	intron 3	35.31704	<.0001***	Abdominal Cuticle
tan	3408	intron 3	47.03838	<.0001***	Abdominal Cuticle
tan	3360	intron 3	31.59926	<.0001***	Abdominal Stripe
tan	3408	intron 3	48.97887	<.0001***	Abdominal Stripe

Table 2.8. List of SNP sites significantly associated with pigmentation, allele states, and significant effect sizes on associated pigmentation traits. Position refers to where the SNP sites resides within the tested gene region starting at the end position of each gene and ending 5kb upstream of the gene. Negative values correspond to lower values for luminosity or darker phenotypes.

Gene	SNP Site	Region	Allele	Thorax Cuticle	Thorax Trident	Abdominal Cuticle	Abdominal Stripe
ebony	312	exon 7	A	--	**-0.2736**	--	--
			G*	--	**0.4495**	--	--
	1671	exon 4	A	**-0.2059**	**-0.2054**	--	--
			G*	**0.5208**	**0.5195**	--	--
	1997	exon 3	A*	**-0.3682**	--	--	--
			G	**0.1882**	--	--	--
	2480	exon 2	A	--	**-0.2186**	--	--
			G*	--	**0.3590**	--	--
	2882	exon 2	A	--	--	--	**-0.2441**
			G*	--	--	--	**0.4177**
	3503	intron 1	G	--	**-0.1271**	--	--
			A*	--	**0.6743**	--	--
	3586	intron 1	G*	**-0.5249**	**-0.7726**	--	--
			A	**0.1167**	**0.1717**	--	--
	3761	intron 1	T*	--	**-0.7036**	--	--
			C	--	**0.1910**	--	--
	3885	intron 1	T*	**-0.5249**	**-0.7726**	--	--
			C	**0.1167**	**0.1717**	--	--
	4130	intron 1	A*	--	**-0.4824**	--	--
			T	--	**0.1848**	--	--
	5134	intron 1	G	--	**-0.2756**	--	--
			A*	--	**0.3064**	--	--
	5825	intron 1	G	--	**-0.3452**	--	--
			A*	--	**0.2855**	--	--
	6182	intron 1	T*	--	**-0.5670**	--	--
			C	--	**0.2149**	--	--
	6245	intron 1	A*	**-0.4853**	**-0.6906**	--	--
			T	**0.1862**	**0.2595**	--	--
	6450	intron 1	A	--	**-0.2246**	--	--
			T*	--	**0.4884**	--	--
	9309	upstream	G	--	--	--	**-0.1107**
			C*	--	--	--	**0.3340**
	9382	upstream	C*	--	--	**-0.9459**	**-1.0933**
			A	--	--	**0.0988**	**0.1141**
	10143	upstream	A*	**-0.3149**	--	--	--

			G	**0.2388**	--	--	--
	10233	upstream	G*	**-0.4017**	**-0.5589**	--	--
			C	**0.1506**	**0.2096**	--	--
	10246	upstream	C	--	**-0.2614**	--	--
			T*	--	**0.4255**	--	--
	10313	upstream	A	**-0.2359**	--	--	--
			T*	**0.5130**	--	--	--
	10525	upstream	A*	--	**-0.5558**	--	--
			G	--	**0.1806**	--	--
	10702	upstream	T*	**-0.4308**	**-0.5926**	--	--
			G	**0.1425**	**0.1918**	--	--
	10935	upstream	A*	--	--	**-0.9459**	**-1.0933**
			T	--	--	**0.0621**	**0.0925**
	10943	upstream	A*	--	--	--	**0.6638**
			T	--	--	--	**-0.1548**
	11116	upstream	G*	**-0.4938**	**-0.6851**	--	--
			T	**0.1170**	**0.1609**	--	--
	11144	upstream	T*	--	--	**-0.9459**	**-1.0933**
			G	--	--	**0.0988**	**0.1141**
	11167	upstream	T	--	--	--	**-0.1408**
			G*	--	--	--	**0.9295**
	11201	upstream	A	--	--	--	**-0.1341**
			T*	--	--	--	**1.1160**
	11397	upstream	A*	--	--	**-0.9459**	**-1.0933**
			G	--	--	**0.0988**	**0.1141**
	11449	upstream	A*	--	--	**-0.9459**	**-1.0933**
			T	--	--	**0.0988**	**0.1141**
	11473	upstream	A*	--	--	**-0.9459**	**-1.0933**
			T	--	--	**0.0988**	**0.1141**
	11484	upstream	T*	--	--	**-0.9459**	**-1.0933**
			A	--	--	**0.0988**	**0.1141**
	11561	upstream	C	**-0.1141**	--	--	**-0.1185**
			T*	**0.7853**	--	--	**0.8219**
pale	1444	intron 4	A	--	--	**-0.1800**	**-0.1504**
			T	--	--	**0.6926**	**0.5934**
	1497	intron 4	T	--	--	**0.3261**	--
			G	--	--	**-0.2891**	--
	7187	upstream	T	--	--	**-0.1780**	--
			G	--	--	**0.5339**	--
	8544	upstream	A	--	**-0.2836**	--	--
			T	--	**0.3676**	--	--
	8568	upstream	A	**-0.2395**	**-0.2836**	--	--
			T	**0.3444**	**0.4079**	--	--

	8587	upstream	G	**-0.2395**	**-0.2836**	--	--
			A	**0.3444**	**0.4079**	--	--
	8615	upstream	A	**-0.2395**	**-0.2836**	--	--
			G	**0.3444**	**0.4079**	--	--
	8635	upstream	T	--	**-0.4249**	--	--
			C	--	**0.2193**	--	--
	8873	upstream	A	--	**-0.4103**	--	--
			G	--	**0.2779**	--	--
	9550	upstream	G	--	**-0.1606**	--	--
			T	--	**0.6198**	--	--
Ddc	1034	exon 2	C	**-0.2573**	--	**-0.2252**	--
			T	**0.4303**	--	**0.3771**	--
	1061	exon 2	T	**-0.1264**	--	**-0.1589**	**-0.1113**
			C	**0.8866**	--	**1.1229**	**0.7869**
	1941	intron 1	A	--	--	--	**-0.3229**
			G	--	--	--	**0.2820**
	2160	intron 1	A	--	**-0.2448**	--	--
			G	--	**0.3515**	--	--
	2784	exon 1	A	--	**-0.6584**	**-0.6504**	**-0.7735**
			G	--	**0.1505**	**0.1496**	**0.1779**
	2797	exon 1	A	--	--	--	**-0.8095**
			G	--	--	--	**0.1490**
	3063	exon 1	C	--	--	**-0.1771**	**-0.1685**
			T	--	--	**0.6256**	**0.5953**
	3371	exon 1	T	--	--	--	**-0.1201**
			G	--	--	--	**0.8324**
	3828	upstream	A	--	--	--	**-0.6923**
			G	--	--	--	**0.1295**
	4338	upstream	T	**-0.3002**	--	--	--
			A	**0.2989**	--	--	--
	4351	upstream	G	**-0.3002**	--	--	--
			C	**0.2989**	--	--	--
	4957	upstream	C	--	--	--	**-0.6888**
			G	--	--	--	**0.0974**
	5440	upstream	G	--	--	**-0.1506**	--
			A	--	--	**0.8207**	--
	5529	upstream	A	**-0.3647**	**-0.3344**	--	--
			G	**0.2814**	**0.2580**	--	--
	6000	upstream	A	--	--	**-0.5265**	**-0.5426**
			T	--	--	**0.0809**	**0.1010**
	8148	upstream	A	**-1.1176**	**-0.9626**	--	--
			C	**0.1118**	**0.0963**	--	--
	8449	upstream	A	**-0.6558**	--	--	--

			G	**0.1197**	--	--	--
	8662	upstream	G	**-0.6161**	--	--	--
			A	**0.1408**	--	--	--
	8721	upstream	A	**-0.9585**	--	--	--
			G	**0.1340**	--	--	--
tan	860	exon 8	T	--	**-0.8726**	--	--
			C	--	**0.1593**	--	--
	2262	intron 3	A	--	--	**-0.1677**	--
			T	--	--	**0.5030**	--
	3360	intron 3	T	--	--	--	**-0.2388**
			G	--	--	--	**0.3301**
	3408	intron 3	C	**-0.2093**	--	**-0.2821**	**-0.2874**
			T	**0.3010**	--	**0.4101**	**0.4178**
	6716	upstream	C	**-0.1011**	--	--	--
			T	**0.7092**	--	--	--
	7939	upstream	G	**-1.0361**	**-1.3399**	--	--
			A	**0.1089**	**0.1409**	--	--
	7945	upstream	A	--	**-0.6580**	--	--
			G	--	**0.1240**	--	--
	8218	upstream	T	--	**0.1740**	--	--
			C	--	**-0.7507**	--	--
	8464	upstream	C	**-1.0361**	**-1.3399**	--	--
			A	**0.1117**	**0.1087**	--	--
	8551	upstream	T	--	**-1.3399**	--	--
			G	--	**0.1087**	--	--
	8996	upstream	A	--	**0.1910**	--	--
			C	--	**-0.6860**	--	--

Table 2.9. List of average negative and positive effect sizes of significant SNPS of each gene on pigmentation traits. Negative values correspond to lower values for luminosity or darker phenotypes.

Gene	Thorax Cuticle	Thorax Trident	Abdominal Cuticle	Abdominal Stripe
ebony	-0.3728	-0.4618	-0.9459	-0.6582
	0.2796	0.3025	0.0935	0.3892
pale	-0.1796	-0.2663	-0.1618	-0.0752
	0.2583	0.3385	0.3881	0.2967
Ddc	-0.5218	-0.5501	-0.3148	-0.4699
	0.3003	0.2141	0.5295	0.3502
tan	-0.4765	-0.8734	-0.1499	-0.1754
	0.2462	0.1258	0.3044	0.2493

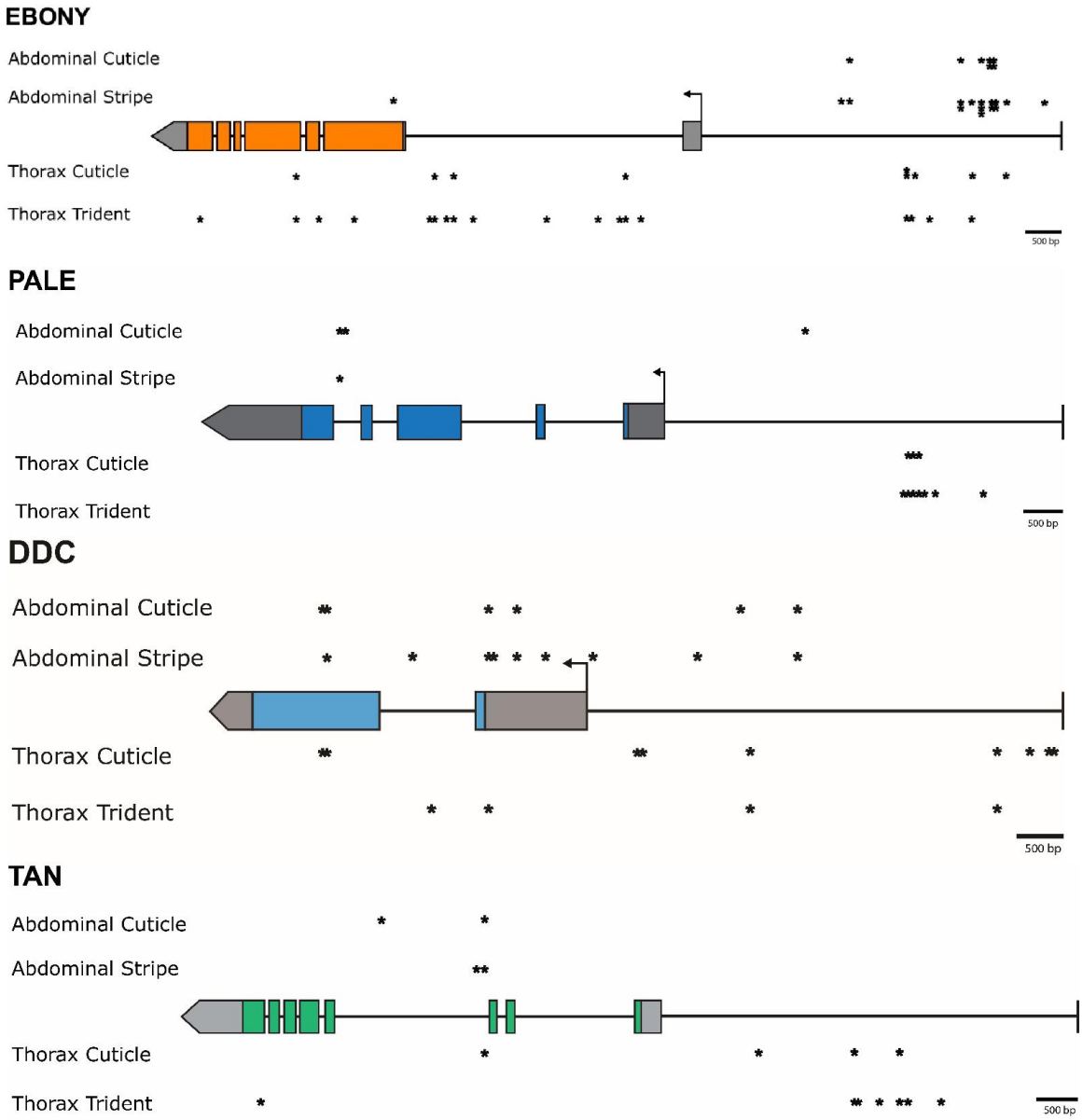


Figure 2.5. Overview of Results of SNP associated study represented by gene models for *ebony*, *pale*, *Ddc* and *tan* with significant SNP sites denoted with stars. Boxes represent exons while lines represent introns and non-coding regions. Arrows denote start codon locations. Colored areas indicate coding regions while gray areas denote non-coding translated regions. Stacked stars indicated multiple SNP sites within the area. Stars line up with text on the left naming associated pigmentation traits. Respective scales are given at the lower right corners of each model.

Table 2.10. List of SNP sites that have significant associations with mRNA expression following the permutation test, and associated effect sizes of alternative allele states. Position refers to where the SNP sites resides within the tested gene region starting at the end position of each gene and ending 5kb upstream of the gene. Position numbers denoted with Δ symbols indicated sites that were also associated with pigmentation traits. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

gene	SNP Site	Location	F	p-value	Allele	Effect Size
ebony	312 Δ	exon 7	11.83	< .01**	A G	-0.196 0.409
	6245 Δ	intron 1	12.84	<.001***	A T	-0.533 0.195
pale	872	exon 5	18.28	<.001***	T G	0.340 -0.290
	1215	exon 5	17.42	<.001***	A G	-0.413 0.224
	1861	intron 3	16.67	<.001***	T C	0.206 -0.429
	2440	exon 3	19.94	<.0001***	T G	0.157 -0.673
	2455	exon 3	19.94	<.0001***	T C	0.157 -0.673
	2476	exon 3	19.94	<.0001***	A T	-0.673 0.157
	2767	exon 3	23.48	<.0001***	T C	0.244 -0.509
	3821	intron 2	20.07	<.0001***	T C	0.254 -0.418
	4110	intron 1	17.87	<.001***	A G	-0.884 0.107
	4844	intron 1	19.85	<.0001***	A T	-0.671 0.157
	5040	exon 1	17.64	<.001***	T G	0.633 -0.148
Ddc	2460	intron 1	9.15	< .01**	T G	-0.486 0.156
tan	109	exon 8	12.97	<.001***	T C	0.326 -0.320
	2574	intron 3	12.59	< .01**	T G	0.708 -0.137

Table 2.11. Summary statistics of sequence diversity (S, π, Θ_w) of the DGRP lines for each candidate gene: within each gene region (introns and exons), upstream regions, and chromosome arm estimates derived from Mackay, Richards et al. 2012. (Loc = Location, Up. = Upstream, Ch. A. = Chromosome Arm)

				S		***π***			***θ$_w$***		
Gene	**Loc.**	**N**	**Gene Length (bp)**	**Gene**	**Up.**	**Gene**	**Up.**	**Ch. A.**	**Gene**	**Up.**	**Ch. A.**
ebony	3R	37	7339	167	192	0.00523	0.00914	0.0051	0.00548	0.00941	0.0063
pale	3L	37	5488	134	118	0.00506	0.00609	0.0061	0.00601	0.00575	0.0074
Ddc	2L	37	3824	74	123	0.00455	0.00701	0.0068	0.00464	0.00589	0.008
tan	X	37	5603	100	101	0.00435	0.00466	0.004	0.00436	0.00484	0.0049

Table 2.12. Summary of Tajima's Test on sequences from DGRP lines for the gene and upstream regions separately. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

			Tajima's *D*	
Gene	**N**	**Gene Length (bp)**	**Gene**	**Upstream**
ebony	37	7339	-0.1745	-0.1056
pale	37	5488	-0.5888	0.2233
Ddc	37	3824	-0.0664	0.7024
tan	37	5603	-0.007	-0.1368

Table 2.13. Summary Statistics of divergence and neutrality tests between coding regions (CDS) from DGRP Lines (N = 37) and consensus *D. simulans* sequences (N=1) using coalescent simulations (10,000 replications). . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Gene	Location	Outgroup	Ingroup	Length	S	Divergence	Tajima's *D*	Fay and Wu's *H*	*E*-Test
ebony	3R	1	37	2640	55	0.026	-0.585	-0.144	-0.387
pale	3L	1	37	1740	45	0.045	-0.358	**-1.566.**	1.190
Ddc	2L	1	37	1533	11	0.016	0.314	**-1.896***	1.910
tan	X	1	37	1164	14	0.031	-0.590	**-2.609***	1.859

Table 2.14. Listing of significant SNP sites within each candidate gene, the gene region in which they are located, associated traits and measures of Tajima's *D* and nucleotide diversity. (TC, Thoracic Cuticle, TT, Thoracic Trident, AC, Abdominal Cuticle, AS, Abdominal Stripe, Gene Expression, EX). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Gene	Region	SNP	Associated Traits	π	θ_w	Tajima's *D*
ebony	*exon 7*	312	TT, EX	0.00256	0.00287	-0.2928
ebony	*exon 4*	1671	TC, TT	0.00659	0.00814	-0.6257
ebony	*exon 3*	1997	TC	0.01029	0.00814	0.638
ebony	*exon2*	2480	TT	0.00482	0.00719	-1.0635
ebony	*exon 2*	2882	AS	0.00416	0.00575	-0.8663
ebony	*intron 1*	3503	TT	0.01102	0.01006	0.3216
ebony	*intron 1*	3586	TC, TT	0.01061	0.00958	0.3606
ebony	*intron 1*	3761	TT	0.00959	0.00862	0.3694
ebony	*intron 1*	3885	TC, TT	0.00882	0.00814	0.2714
ebony	*intron 1*	4130	TT	0.0045	0.00527	-0.4476
ebony	*intron 1*	5134	TT	0.00384	0.00287	0.9099
ebony	*intron 1*	5825	TT	0.00502	0.00479	0.1455
ebony	*intron 1*	6182	TT	0.00375	0.00479	-0.6596
ebony	*intron 1*	6245	TC, TT, EX	0.0043	0.00479	-0.3102
ebony	*intron 1*	6450	TT	0.00499	0.00527	-0.1631
ebony	*upstream*	9309	AS	0.0097	0.01198	-0.7628
ebony	*upstream*	9382	AC, AS	0.00848	0.01054	-0.7848
ebony	*upstream*	10143	TC	0.01022	0.01389	-0.92
ebony	*upstream*	10233	TC, TT	0.01055	0.01294	-0.638
ebony	*upstream*	10246	TT	0.00975	0.0115	-0.5202
ebony	*upstream*	10313	TC	0.00864	0.00671	0.921
ebony	*upstream*	10525	TT	0.00486	0.00623	-0.6962
ebony	*upstream*	10702	TC, TT	0.00927	0.00958	-0.108
ebony	*upstream*	10935	AC, AS	0.01547	0.01581	-0.0765
ebony	*upstream*	10943	AS	0.01718	0.01677	0.0873
ebony	*upstream*	11116	TC, TT	0.01728	0.01677	0.0065
ebony	*upstream*	11144	AC, AS	0.01625	0.01581	-0.0093
ebony	*upstream*	11167	AS	0.01635	0.01629	-0.087
ebony	*upstream*	11201	AS	0.01633	0.01581	0.0089
ebony	*upstream*	11397	AC, AS	0.01201	0.01102	0.1527
ebony	*upstream*	11449	AC, AS	0.00995	0.00958	-0.0365
ebony	*upstream*	11473	AC, AS	0.01015	0.00958	0.1988
ebony	*upstream*	11484	AC, AS	0.01015	0.00958	0.1988
ebony	*upstream*	11561	TC, AS	0.00951	0.00958	-0.026
pale	*intron 4*	1444	AC, AS	0.01365	0.01389	-0.1749

pale	*intron 4*	1497	AC	0.01243	0.01341	-0.3671
pale	*upstream*	7187	AC	0.00906	0.00623	0.8417
pale	*upstream*	8544	TT	0.01423	0.01102	0.9909
pale	*upstream*	8568	TC, TT	0.01423	0.01102	0.9909
pale	*upstream*	8587	TC, TT	0.01423	0.01102	0.9909
pale	*upstream*	8615	TC, TT	0.01423	0.01102	0.9909
pale	*upstream*	8635	TT	0.01524	0.0115	1.1111
pale	*upstream*	8873	TT	0.00936	0.00814	0.492
pale	*upstream*	9550	TT	0.00492	0.00623	-0.6626
ddc	*exon 2*	1034	TC, AC	0.00298	0.00299	-0.012
ddc	*exon2*	1061	TC, AC, AS	0.00312	0.00359	-0.3609
ddc	*intron 1*	1941	AS	0.00612	0.00659	-0.2193
ddc	*intron 1*	2160	TT	0.00584	0.00779	-0.7908
ddc	*exon 1*	2784	TT, AC, AS	0.00584	0.00779	-0.7908
ddc	*exon 1*	2797	AS	0.00584	0.00779	-0.7908
ddc	*exon 1*	3063	AC, AS	0.00531	0.00659	-0.5985
ddc	*exon 1*	3371	AS	0.00405	0.00539	-0.7359
ddc	*upstream*	3828	AS	0.00867	0.00599	1.3569
ddc	*upstream*	4338	TC	0.00878	0.00659	1.0272
ddc	*upstream*	4351	TC	0.00818	0.00599	1.1101
ddc	*upstream*	4957	AS	0.01036	0.01138	-0.2982
ddc	*upstream*	5440	AC	0.00824	0.00898	-0.2663
ddc	*upstream*	5529	TC, TT	0.00724	0.00838	-0.4351
ddc	*upstream*	6000	AC, AS	0.00346	0.00419	-0.4911
ddc	*upstream*	8148	TC, TT	0.00077	0.0018	-1.2758
ddc	*upstream*	8449	TC	0.00459	0.00419	0.2704
ddc	*upstream*	8662	TC	0.00342	0.0025	0.8992
ddc	*upstream*	8721	TC	0.00342	0.0025	0.8992
tan	*exon 8*	860	TT	0.00164	0.0024	-0.8175
tan	*intron 3*	2262	AC	0.003	0.0024	0.1213
tan	*intron 3*	3360	AS	0.00439	0.00479	-0.2533
tan	*intron 3*	3408	TC, AC, AS	0.00515	0.00527	-0.0718
tan	*upstream*	6716	TC	0.00374	0.00431	-0.3967
tan	*upstream*	7939	TC, TT	0.00569	0.00479	0.567
tan	*upstream*	7945	TT	0.00724	0.00575	0.8125
tan	*upstream*	8218	TT	0.00323	0.00383	-0.4595
tan	*upstream*	8464	TC, TT	0.00295	0.00287	0.0759
tan	*upstream*	8551	TT	0.00377	0.00335	0.346
tan	*upstream*	8996	TT	0.00444	0.00383	0.4582

Table 2.15. Summary of results of McDonald Kreitman (MK) Test using coding regions for pigmentation gene sequences from DGRP lines (N=37) and White 105 *D. simulans* sequence (N = 1). NI (Neutrality Index), α, and results from Fisher's exact test and G (goodness of fit test using William's correction) are also reported. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

a. MK results for the *ebony* gene

	Diverged	Polymorphic	Total	NI	α	Fisher's Exact Test	G
Synonymous	100	44	144	5.909	-4.909	**<.01****	**11.25*****
Nonsynonymous	5	13	18				
Total	105	57	162				

b. MK results for the *pale* gene

	Diverged	Polymorphic	Total	NI	α	Fisher's Exact Test	G
Synonymous	10	42	52	0.397	0.603	0.353	1.105
Nonsynonymous	3	5	8				
Total	13	47	60				

c. MK results for the *Ddc* gene

	Diverged	Polymorphic	Total	NI	α	Fisher's Exact Test	G
Synonymous	34	10	44	6.800	-5.800	0.1560	1.954
Nonsynonymous	1	2	3				
Total	35	12	47				

d. MK results for the *tan* gene

	Diverged	Polymorphic	Total	NI	α	Fisher's Exact Test	G
Synonymous	26	18	44	0.000	1.000	0.353	N/A
Nonsynonymous	3	0	3				
Total	29	18	47				

Table 2.16. Results of the HKA test for the coding region of each candidate gene against a consensus sequence in *D. simulans*. This test was conducting using coalescent simulations to estimate population parameters and expected values of θ under neutral expectations. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Gene		Observed	Expected	Variance	SD	
ebony	Polymorphisms	53	62.38	423.6	0.208	
	Divergence	113.05	103.67	324.96	0.271	
tan	Polymorphisms	17	16.71	42.62	0.002	
	Divergence	35.57	35.86	51.88	0.002	
Ddc	Polymorphisms	10	18.02	48.17	1.336	
	Divergence	38.11	30.09	48.72	1.321	
pale	Polymorphisms	42	24.89	82.37	**3.556**	**
	Divergence	24.43	41.55	77.08	**3.8**	**

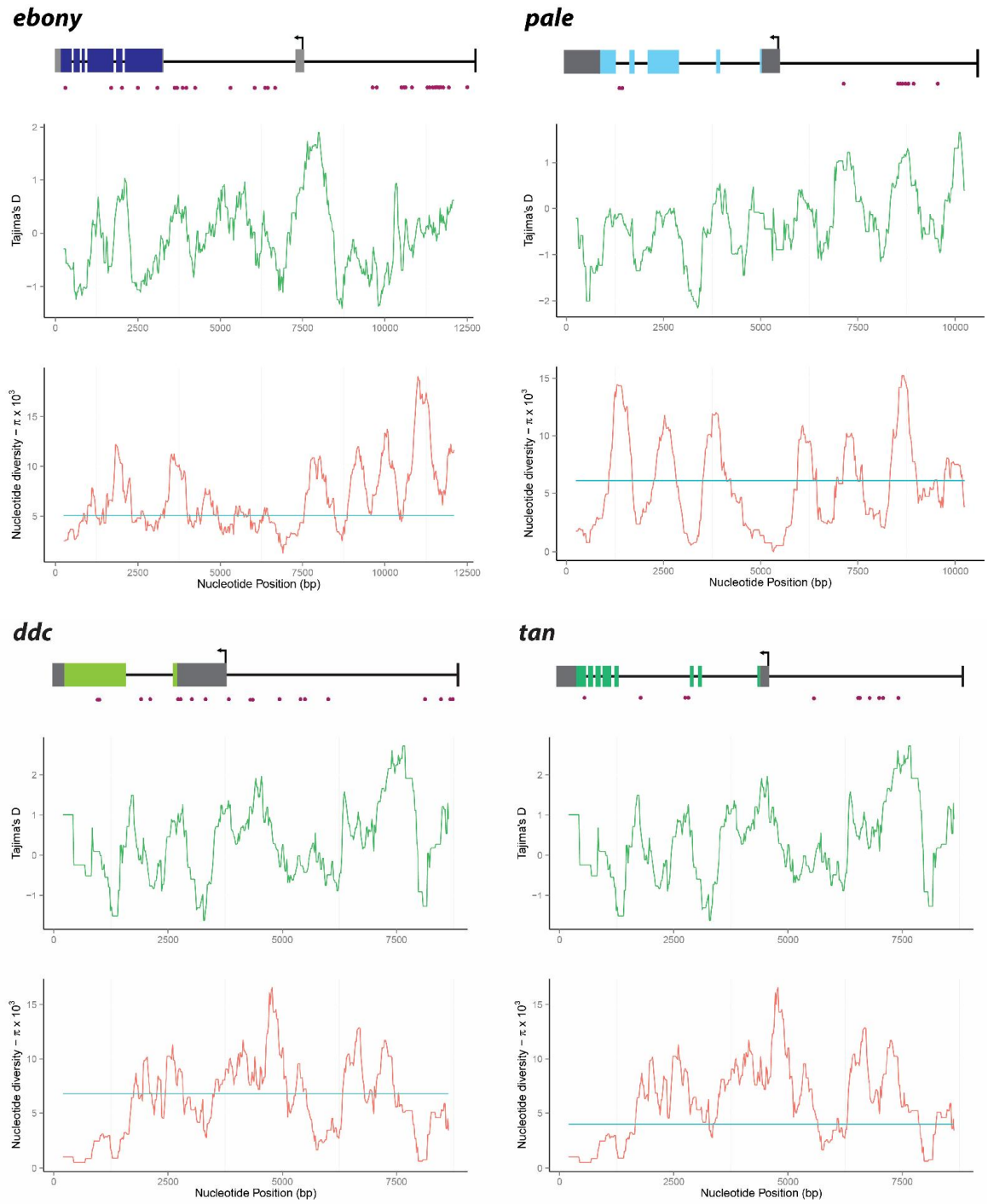


Figure 2.6. Estimates of Tajimas *D* and Nucleotide Diversity (π) across each candidate gene and associated regions upstream of the start codon (5kb). In each the gene model, Boxes represent exons while lines represent introns and non-coding regions. Arrows denote start codon locations. Colored areas indicate coding regions while gray areas denote non-coding translated regions. Blue lines indicate estimates of π for the genome region (10 kb upstream and downstream of each gene).

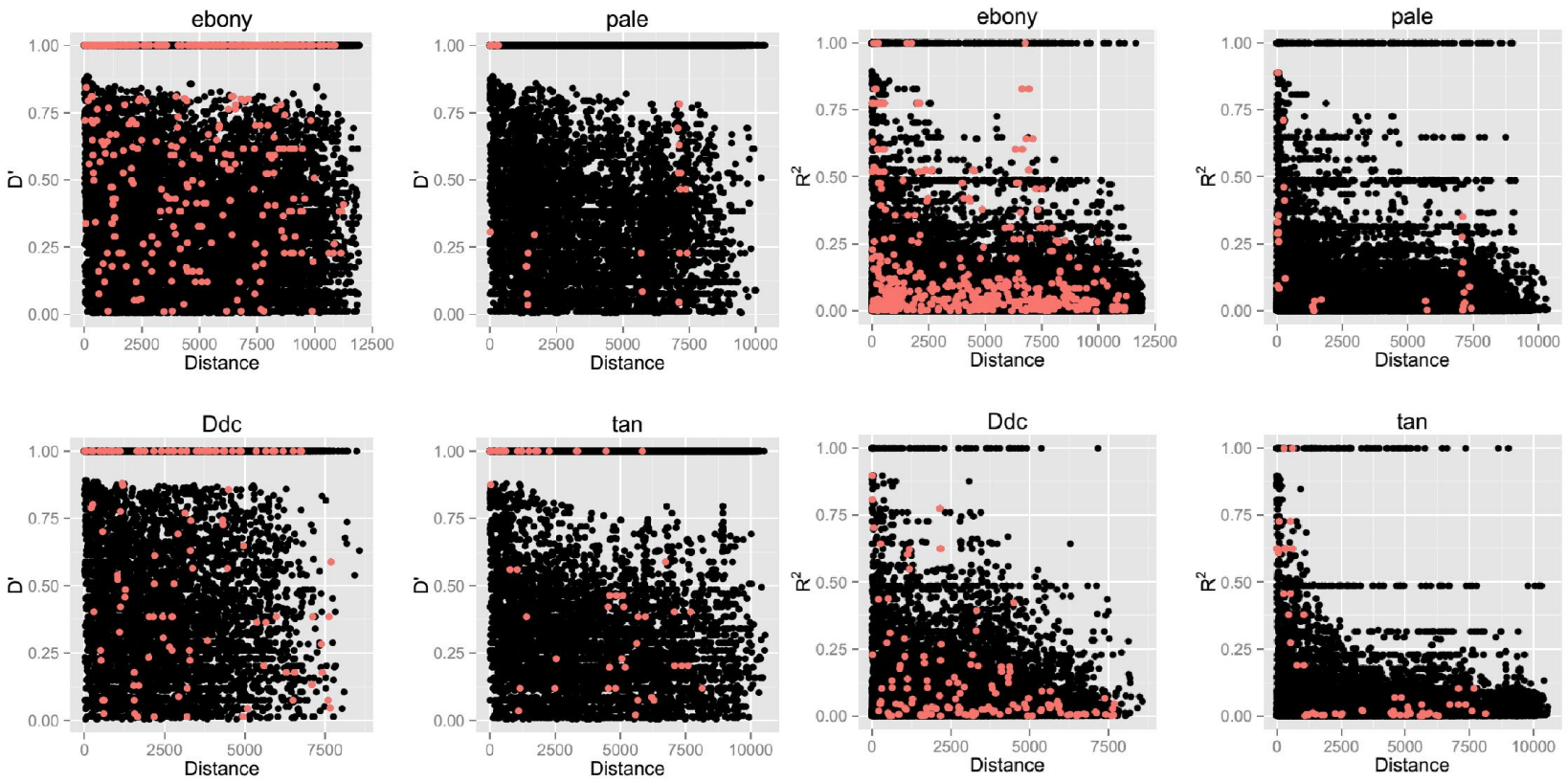


Figure 2.7. Estimates of Linkage Disequilibrium between polymorphic sites within candidate genes using *D'* and R^2 values. Each point represents pairwise comparisons of Linkage Disequilibrium estimates between two SNP sites. Pink dots denote comparisons among SNP sites that are significantly associated with pigmentation traits.

Figure 2.8. LD as illustrated as a matrix of pairwise comparisons of estimates of R^2 values among SNP sites associated with pigmentation traits in the candidate genes (a. *ebony*, b.*pale*, c. *Ddc*, and d. *tan*). The locations of SNP markers are represented in red text. *P*-values are denoted by blue (significant after Bonferroni correction), green, ($P < 0.001$), yellow, ($0.001 < P < 0.01$) or grey ($0.01 < P < 0.05$). Significance was determine using a X^2 tests for each comparison.

a. LD (R^2) matrix for *ebony*

	312	1671	1997	2480	2882	3503	3586	3761	3885	4130	5134	5825	6182	6245	6450	9382	10143	10233	10246	10313	10525	10702	10935	10943	11116	11144	11167	11201	11397	11449	11473	11484	11561
312		0.486	0.072	0.001	0.027	0.027	0.093	0.112	0.093	0.000	0.009	0.045	0.085	0.112	0.314	0.003	0.105	0.154	0.030	0.257	0.112	0.093	0.003	0.013	0.042	0.003	0.017	0.058	0.011	0.003	0.003	0.011	0.054
1671			0.127	0.004	0.026	0.026	0.045	0.054	0.045	0.002	0.000	0.018	0.019	0.054	0.194	0.028	0.099	0.075	0.037	0.065	0.054	0.045	0.028	0.007	0.020	0.028	0.028	0.007	0.036	0.028	0.028	0.036	0.000
1997				0.035	0.105	0.019	0.358	0.262	0.358	0.140	0.171	0.025	0.100	0.135	0.053	0.085	0.262	0.256	0.116	0.151	0.262	0.197	0.085	0.015	0.163	0.085	0.005	0.015	0.042	0.012	0.085	0.042	0.016
2480					0.029	0.358	0.105	0.045	0.105	0.174	0.025	0.090	0.037	0.045	0.070	0.085	0.000	0.023	0.011	0.047	0.045	0.029	0.085	0.051	0.048	0.085	0.012	0.015	0.042	0.012	0.085	0.042	0.016
2882						0.038	0.038	0.001	0.038	0.062	0.004	0.008	0.011	0.045	0.016	0.023	0.082	0.062	0.069	0.027	0.001	0.038	0.023	0.005	0.017	0.023	0.102	0.144	0.030	0.023	0.023	0.030	0.065
3503							0.038	0.045	0.038	0.062	0.008	0.056	0.072	0.045	0.038	0.308	0.001	0.062	0.002	0.000	0.045	0.038	0.308	0.005	0.017	0.308	0.023	0.005	0.221	0.102	0.308	0.221	0.030
3586								0.830	1.000	0.602	0.147	0.127	0.523	0.524	0.082	0.023	0.266	0.602	0.118	0.093	0.830	0.642	0.023	0.005	0.456	0.023	0.023	0.005	0.030	0.023	0.023	0.030	0.030
3761									0.830	0.477	0.080	0.076	0.407	0.420	0.099	0.028	0.194	0.477	0.142	0.035	1.000	0.524	0.028	0.007	0.378	0.028	0.003	0.007	0.036	0.028	0.028	0.036	0.000
3885										0.602	0.147	0.127	0.523	0.524	0.082	0.023	0.266	0.602	0.118	0.093	0.830	0.642	0.023	0.005	0.456	0.023	0.023	0.005	0.030	0.023	0.023	0.030	0.030
4130											0.013	0.156	0.256	0.281	0.002	0.039	0.102	0.314	0.195	0.000	0.477	0.366	0.039	0.009	0.275	0.039	0.039	0.009	0.050	0.039	0.039	0.050	0.002
5134												0.187	0.166	0.080	0.072	0.092	0.108	0.058	0.011	0.108	0.080	0.056	0.092	0.021	0.067	0.092	0.050	0.036	0.034	0.016	0.016	0.034	0.018
5825													0.204	0.076	0.008	0.092	0.001	0.072	0.014	0.045	0.076	0.043	0.092	0.036	0.020	0.092	0.002	0.036	0.119	0.092	0.092	0.119	0.086
6182														0.630	0.157	0.045	0.162	0.420	0.122	0.026	0.407	0.523	0.045	0.010	0.238	0.045	0.045	0.010	0.058	0.045	0.045	0.058	0.058
6245															0.099	0.028	0.194	0.477	0.055	0.112	0.420	0.524	0.028	0.007	0.378	0.028	0.028	0.007	0.036	0.028	0.028	0.036	0.036
6450																0.001	0.027	0.053	0.010	0.094	0.099	0.082	0.001	0.066	0.037	0.001	0.001	0.012	0.007	0.001	0.001	0.007	0.007
9382																	0.119	0.039	0.074	0.058	0.028	0.023	1.000	0.003	0.011	1.000	0.015	0.003	0.776	0.518	0.518	0.776	0.019
10143																		0.539	0.257	0.203	0.194	0.266	0.119	0.066	0.209	0.119	0.051	0.012	0.069	0.024	0.024	0.069	0.066
10233																			0.195	0.154	0.477	0.602	0.039	0.086	0.275	0.039	0.039	0.009	0.050	0.039	0.039	0.050	0.050
10246																				0.033	0.142	0.118	0.074	0.017	0.054	0.074	0.071	0.046	0.021	0.008	0.008	0.021	0.033
10313																					0.035	0.093	0.058	0.013	0.042	0.058	0.100	0.058	0.075	0.058	0.058	0.075	0.054
10525																						0.524	0.028	0.007	0.378	0.028	0.003	0.007	0.036	0.028	0.028	0.036	0.000
10702																							0.023	0.005	0.456	0.023	0.023	0.005	0.030	0.023	0.023	0.030	0.030
10935																								0.003	0.011	1.000	0.015	0.003	0.776	0.518	0.518	0.776	0.019
10943																									0.003	0.003	0.003	0.001	0.004	0.003	0.003	0.004	0.004
11116																										0.011	0.011	0.003	0.014	0.011	0.011	0.014	0.014
11144																											0.015	0.003	0.776	0.518	0.518	0.776	0.019
11167																												0.229	0.019	0.015	0.015	0.019	0.392
11201																													0.004	0.003	0.003	0.004	0.178
11397																														0.776	0.776	1.000	0.024
11449																															0.518	0.776	0.019
11473																																0.776	0.019
11484																																	0.024
11561																																	

b. LD (R^2) matrix for *pale*

1444	0.358	0.004	0.002	0.002	0.019	0.069	0.029	0.008
1497	0.036	0.138	0.138	0.277	0.352	0.179	0.090	
7187	0.013	0.013	0.004	0.001	0.036	0.042		
8544	0.094	0.288	0.257	0.085	0.119			
8568	0.288	0.257	0.085	0.119				
8587	0.889	0.294	0.412					
8615	0.329	0.464						
8635	0.711							
8873								

c. **LD (R^2) matrix for *Ddc***

1034	0.229	0.041	0.602	0.132	0.106	0.028	0.027	0.318	0.394	0.082	0.424	0.067	0.012	0.004	0.028	0.046
1061	0.178	0.025	0.030	0.024	0.031	0.024	0.012	0.008	0.019	0.042	0.000	0.014	0.065	0.000	0.002	
1941	0.019	0.005	0.004	0.100	0.004	0.033	0.029	0.003	0.024	0.007	0.003	0.005	0.008	0.005		
2160	0.132	0.106	0.188	0.000	0.206	0.268	0.082	0.185	0.014	0.012	0.004	0.001	0.004			
2784	0.806	0.053	0.221	0.001	0.000	0.626	0.033	0.001	0.023	0.038	0.053	0.038				
2797	0.043	0.289	0.002	0.005	0.776	0.012	0.000	0.019	0.030	0.043	0.030					
3063	0.043	0.049	0.062	0.033	0.030	0.007	0.001	0.092	0.041	0.092						
3828	0.012	0.008	0.138	0.002	0.000	0.019	0.030	0.043	0.030							
4338	0.897	0.001	0.549	0.198	0.103	0.165	0.124	0.165								
4351	0.000	0.623	0.221	0.115	0.183	0.144	0.183									
4957	0.001	0.028	0.015	0.023	0.033	0.023										
5529	0.275	0.143	0.109	0.192	0.228											
6000	0.076	0.122	0.062	0.122												
8148	0.102	0.440	0.308													
8449	0.436	0.642														
8662	0.702															
8721																

d. LD (R^2) matrix for *tan*

860	0.011	0.002	0.008	0.005	0.102	0.042	0.042	0.019	0.102	0.008
2262	0.001	0.007	0.010	0.000	0.011	0.011	0.002	0.000	0.041	
3360	0.610	0.017	0.008	0.069	0.069	0.031	0.008	0.043		
3408	0.021	0.016	0.004	0.004	0.003	0.016	0.000			
6716	0.003	0.005	0.005	0.003	0.003	0.009				
7939	0.626	0.626	0.728	1.000	0.377					
7945	1.000	0.456	0.626	0.188						
8218	0.456	0.626	0.188							
8464	0.728	0.275								
8551	0.377									
8996										

Table 2.17. Results from multivariate Linear Regression Models of expression of pigmentation genes on female pigmentation. . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Coefficients	**Abdominal Cuticle**			**Abdominal Stripe**			**Thoracic Cuticle**			**Thoracic Trident**		
	Estimate	SE		Estimate	SE		Estimate	SE		Estimate	SE	
(Intercept)	1.002	**0.1852**	***	0.2763	**-0.1049**	*	0.39880000	**0.1071**	***	0.40260000	**0.1334**	**
ebony	-0.001988	**0.0007018**	**	-0.0007028	**-0.0003977**	.	-0.00045390	0.000406		-0.00050320	0.000506	
pale	-0.01022	**0.004325**	*	-0.002344	-0.002451		-0.00257200	0.002501		-0.00418600	0.003116	
tan	-0.00005023	0.00006728		-0.00000338	-0.00003813		0.00000657	3.89E-05		-0.00002826	4.85E-05	
Ddc	-0.00002147	0.000176		0.00008386	-0.00009973		-0.00000130	0.000102		-0.00008447	0.000127	
ebony x pale	0.000039	**0.0000174**	*	0.00000945	-0.000009859		0.00000958	$1.01X\ 10^{-5}$		0.00001679	1.2510^{-5}	
Adjusted R^2	0.2082			0.0389			-0.1155			-0.09448		
N	32			32			32			32		

Table 2.18. Chromosomal and cytological locations of candidate genes and list of possible overlapping inversions identified in Huang, Massouras et al. 2014.

Gene	Chromosomal Location	Cytological Location	Possible Overlapping Inversions
ebony	3R	93C7 - 93D1	*In(3R)P, In(3R)K, In(3R)Mo, In(3R)C*
pale	3L	65C3	*In(3L)P, In(3L)M, In(3L)Y*
Ddc	2L	37C1	none
tan	X	8D1	none

Table 2.19. Cytological analysis of polymorphic inversions that may overlap with candidate genes derived from Huang, Massouras et al. 2014. Lines included those used in present study. No inversion data was available for line DGRP_514. IN = homozygous for the inversion; ST = homozygous for the standard arrangement; IN/ST heterozygous for the inversion.

DGRP Line	***In(3L)P***	***In(3L)M***	***In(3L)Y***	***In(3R)P***	***In(3R)K***	***In(3R)Mo***	***In(3R)C***
DGRP_301	ST	ST	ST	ST	ST	ST	ST
DGRP_303	ST	ST	ST	ST	ST	ST	ST
DGRP_304	ST	ST	ST	ST	ST	ST	ST
DGRP_306	ST	ST	ST	ST	ST	ST	ST
DGRP_307	ST	ST	ST	ST	ST	ST	ST
DGRP_313	ST	ST	ST	ST	ST	ST	ST
DGRP_315	ST	ST	ST	ST	ST	ST	ST
DGRP_324	ST	ST	ST	ST	ST	**INV**	ST
DGRP_335	ST	ST	ST	ST	ST	**INV/ST**	ST
DGRP_357	ST	ST	ST	ST	ST	ST	ST
DGRP_358	ST	ST	ST	ST	ST	**INV**	ST
DGRP_360	ST	ST	ST	ST	ST	ST	ST
DGRP_362	ST	ST	ST	ST	ST	ST	ST
DGRP_365	ST	ST	ST	ST	ST	ST	ST
DGRP_375	ST	ST	ST	ST	ST	ST	ST
DGRP_379	ST	ST	ST	ST	ST	ST	ST
DGRP_380	ST	ST	ST	ST	ST	ST	ST
DGRP_391	ST	ST	ST	ST	ST	ST	ST
DGRP_399	ST	ST	ST	ST	ST	ST	ST
DGRP_427	ST	ST	ST	ST	ST	ST	ST
DGRP_437	ST	ST	ST	ST	ST	**INV**	ST
DGRP_486	ST	ST	ST	ST	ST	ST	ST
DGRP_514	N/A	N/A	N/A	N/A	N/A	N/A	N/A
DGRP_517	ST	ST	ST	ST	ST	ST	ST

DGRP_555	ST	ST	ST	ST	ST	**INV**	ST
DGRP_639	ST	ST	ST	ST	ST	ST	ST
DGRP_705	ST	ST	ST	ST	ST	ST	ST
DGRP_707	ST	ST	ST	ST	ST	**INV**	ST
DGRP_714	ST	ST	ST	ST	ST	**INV**	ST
DGRP_730	ST	ST	ST	ST	ST	ST	ST
DGRP_732	ST	ST	ST	ST	**INV/ST**	ST	ST
DGRP_765	ST	ST	ST	ST	ST	ST	ST
DGRP_774	ST	ST	ST	ST	ST	ST	ST
DGRP_786	ST	ST	ST	**INV**	ST	ST	ST
DGRP_799	ST	ST	ST	ST	ST	ST	ST
DGRP_820	ST	ST	ST	ST	ST	**INV**	ST
DGRP_852	ST	ST	ST	ST	ST	ST	ST

Chapter 3. Geographic variation in pigmentation traits and genes North American populations of *Drosophila melanogaster*

Introduction

Examining how phenotypes vary within and among populations is informative in understanding their evolution. Geographic patterns may also suggest a possible adaptive component to a trait, in adaptation to environmental gradients or selective pressures that have spatial patterns. In addition to measuring phenotypic differences, one can observe how alleles in contributing genes may be segregating in these populations. Animal pigmentation is of particular interest since there is often a great deal of variation within and among populations. This has been documented in many animal species such as the freshwater isopod *Asellus aquatic*us (Hargeby, Johansson et al. 2004), the beach mouse *Peromyscus polionotus* (Hoekstra, Hirschmann et al. 2006), and the desert lizard species *Sceloporus undulatus*, and *Aspidoscelis inornata* (Hargeby, Johansson et al. 2004). In all these cases variation in the intensity of pigmentation contributes to overall body color and different color morphs tend to show geographic patterns associated with factors affecting adaptation.

Insect species have become of increasing interest since they often show striking diversity and geographic patterns in pigmentation patterns. For example, within species of *Colias* butterflies variation in wing melanization in seen in populations associated with different climates. Heavier melanized wings and undersides are associated with

individuals in regions from higher altitudes or latitudes. Melanin may be playing a thermoregulatory role in this system as darker individuals were shown to absorb heat faster and retain a higher body temperature longer than lighter individuals (Watt 1968). The potential for melanin to contribute to thermoregulation in insects has also been demonstrated in *Drosophila elegans*. Black-morphs were on average .26°C higher than brown-morphs after experimental light irradiation (Hirai and Kimura 1997). In *Drosophila* species, geographic patterns in pigmentation differences have been demonstrated in *D. polymorpha* (Brisson, Toni et al. 2005), *D. americana* (Wittkopp, Stewart et al. 2009), and *D. melanogaster* (Munjal, Karan et al. 1997, Pool and Aquadro 2007). *Drosophila melanogaster* has been well studied in this regard due to the wide breadth of genetic tools available in this model system. In this species, variation in pigmentation traits has been demonstrated among naturally occurring populations throughout the world. Variation is seen in the intensity of the thoracic trident pigmentation pattern on the dorsal side of the thorax (David, Capy et al. 1985, Takahashi, Takahashi et al. 2007), as well as abdominal stripe patterning (Kopp, Graze et al. 2003).

Notable geographic patterns in pigmentation in *D. melanogaster* have been identified with respect to altitude and latitude. Altitudinal clines in abdominal pigmentation have been identified in populations in sub-Saharan Africa (Pool and Aquadro 2007). Similarly, latitudinal differences in pigmentation intensity have been identified between India highland and lowland populations (Munjal, Karan et al. 1997) and among Australian populations (Telonis-Scott, Hoffmann et al. 2011). In all cases,

increases in altitude or latitude are correlated with increases in pigmentation levels. Desiccation tolerance, which is positively associated with pigmentation, also increases with higher latitudes and provides a possible selective mechanism (Kalmus 1941, Parkash, Rajpurohit et al. 2008, Rajpurohit, Parkash et al. 2008). These patterns suggest that environmental factors such as climate may provide the selective agent that results in the observed patterns of pigmentation variation. Other potentially adaptive traits have been demonstrated to exhibit latitudinal patterns in *D. melanogaster* such as thermal sensitivity (Hoffmann, Anderson et al. 2002), thoracic size, and larval developmental time in Australian populations (James, Azevedo et al. 1995), and the ability to diapause in North American populations. (Schmidt and Paaby 2008, Schmidt, Zhu et al. 2008).

In a few cases, the genetic basis of geographic variation in pigmentation traits in *Drosophila* has been identified. In *D. americana*, morphs in the eastern United States are typically darker than western morphs. This variation was associated with DNA sequence variation in the *tan* and *ebony* genes. These loci were also implicated in divergence in body color between *D. americana* and its sister species *D. novamexicana* (Wittkopp, Stewart et al. 2009). Pigmentation differences in Sub-Saharan populations of *D. melanogaster* were associated with the upstream region of *ebony,* which showed strong haplotype structure in the darkest lines. The *tan* and *ebony* genes are both involved in the biosynthesis pathway that produces pigment molecules in the developing fly cuticle (Wright 1987). This pathway is reported to be conserved among *Drosophila* species and other insects which suggest that these genes may serve similar functions in

different species (Blenau and Baumann 2005, Futahashi, Sato et al. 2008, Wittkopp and Beldade 2009, Miyazaki, Okada et al. 2014).

My previous work has revealed that single nucleotide polymorphisms (SNPs) in both *ebony* and *pale* are significantly associated with variation in pigmentation traits in lines representative of a population from Raleigh, North Carolina (see Chapter 2). These lines were developed as part of the Drosophila Genome Reference Panel (DGRP) and are accompanied by mostly complete genome sequence data. Some SNPs were of particular interest since they were clustered and showed possible balancing or directional selection. Since these two candidate genes are both located on the third chromosome, extraction strains containing whole third chromosomes on a uniform X and second chromosome background were generated from strains that are representative of five populations of *Drosophila melanogaster* in Eastern United States. Phenotypic data were collected from these lines to determine the contribution of the third chromosome to natural variation in thoracic and abdominal pigmentation, and also to examine geographic patterns that may exist in these traits. Possible regulatory regions containing SNPs significantly associated with pigmentation variation in the Raleigh population were sequenced to determine whether these sites were correlated with pigmentation in the other populations and/or show geographic patterns of variation. The sequenced region in pale includes six candidate SNP sites that are located 3057 bp upstream of the start codon. The sequenced region in *ebony* includes nine candidate SNP sites that are located 3604 bp upstream of the start codon. This region has already been identified as containing an enhancer for abdominal pigmentation (Rebeiz, Pool et al. 2009). Allele frequencies in

each candidate site within the populations were analyzed to determine if there are latitudinal patterns of variation.

Materials and Methods

Isofemale lines derived from five populations representative of locations throughout Eastern United States were studied: Bowdoin, Maine (BME), Media, Pennsylvania (MPA), Raleigh, North Carolina (RNC), Jacksonville, Florida (JFL), and Homestead, Florida (HFL). RNC lines were obtained from *Drosophila* Genome Research Panel (DGRP, http://dgrp.gnets.ncsu.edu/). These populations are spaced approximately five degrees apart in latitude. Third chromosomes from isofemale lines from each population were extracted into an isogenic background utilizing the *TM2 Ubx*, *TM6 Tb*, and *CyO* balancer chromosomes. The second chromosome of the uniform background was derived from the isogenic laboratory strain 6326 from the Indiana University Bloomington stock center. The X chromosome from the uniform background was derived from the *white*^{1118} strain (Bloomington stock center). A series of four crosses was carried out to complete the extraction and produce the following genotype: w^{1118}/w^{1118}; 6326/6326; extracted 3rd/extracted 3rd; (Figure 3.1). Individuals that were homozygous for the extracted third chromosome were assayed for pigmentation. About 10 - 25 lines were assayed for each population. This depended on the ability to obtain viable homozygous individuals in each line.

Phenotyping assays

Five cultures of five males and five females from each line were set up and kept on standard yeast medium at 21˚C. After four to five days the parents were removed. Newly emerging homozygous adult females were collected from the three most prolific

vials, and aged in fresh food vials for five to seven days. Flies were individually photographed under 50X magnification using an AxioCam (Zeiss) video camera attached to a Leica MZ7 microscope. AxioVision software (Rel. 4.3) was used to capture the images using uniform settings and lighting. Measured regions included those within and outside the thoracic trident, and within and outside the stripe present on the A4 abdominal segment (Figure 3.2). These regions were chosen based on consistent visibility and absence of glare and bristle obstruction. Luminosity, which measures the brightness of selected pixels in an image, was recorded in areas of interest using Image J software (Schneider, Rasband et al. 2012).

Nested ANOVAs were used to determine if there were significant differences among the extraction lines and the five populations in thoracic and abdominal luminosity measurements. This would be indicative of a contribution of genetic factors on the third chromosome to variation in pigmentation phenotypes as well as geographic patterns in these traits. A multi-level linear regression was used to test if any geographic variation may be clinal among the five tested populations. Previously, some of the DGRP isogenic lines were assayed for pigmentation. Eleven of these lines have been successfully used to generate third chromosome extraction lines. For each measured pigmentation trait, correlations coefficient were calculated between the DGRP lines and the extraction lines to see if there is a detectable effect of the third chromosome on any of the pigmentation traits. This would allow us to gain a better idea of to what degree genetic variation on the third chromosome affects each trait and how the genetic background may be influencing phenotypic expression.

DNA Sequencing

Homozygous individuals were collected from the assayed extraction lines. In some lines the yields were low and the number of represented lines varied along each population. Genomic DNA was extracted from single flies representative of phenotyped extraction lines derived from the BME, MPA, RNC, JFL, and HFL populations. Flies were homogenized in a 250 µl of solution of Proteinase K and HB buffer and incubated at 55ºC overnight. Eighty-five µl of 5M NaCl was added to each sample and then the tube was spun down for 20 minutes. The supernatant was transferred to a new tube and 340µl of 100% EtOH was added. Each sample kept in -20ºC overnight. Samples were then spun for 20 minutes and the EtOH was replaced with 70% EtOH. They were spun for 15 minutes, EtOH was removed and each sample was air dried and then re-suspended in 25 µl of water.

For sequencing, we decided to target regions in *ebony* and *pale* containing some of the candidate alleles uncovered from the SNP association study on the DRGP lines. Candidate regions were identified upstream of the start codon for each gene (Figure 3.3). Primers were designed using Primer 3 software and obtained from Sigma-Aldrich Co. (http://biotools.umassmed.edu/bioapps/primer3_www.cgi). The reactions for the *pale* fragment, 483 bp long (primers 5' TGA TGT TGA GAA TAA GAA TAG AAA GCA 3', 5' AGG CTG CAA GAT CGT GTG AT 3'), were annealed at 48˚C in a $MgCl_2$ concentration of 0.25 µmol. The reactions for the *ebony* fragment, 746 bp long (primers 5' GGA TTC GAT TGT AGC CCA GA 3', 5' GTG TGG CTG CAA CTT GTC AC 3'), were annealed at

48˚C in a $MgCl_2$ concentration of 0.25 µmol. Both reactions used 40 cycles of amplification.

Analysis of sequence data

Candidate regions upstream of *pale* and *ebony* were sequenced for this study (Figure 3.3). PCR product cleanup, quantification, and sequencing were performed at the University of Arizona Genetics Core (http://uagc.arl.arizona.edu/) Polymorphic sites were identified in each of the sequences and tested to determine if there is an association with pigmentation traits. The location effect for each gene was removed and the residuals were standardized using the total sample variance. A nested analysis of variance (ANOVA) was used to conduct a SNP association test between candidate SNPs and the standardized residuals of the pigmentation phenotype scores. In order to determine if any of the SNPs had latitudinal patterns, allele frequencies for each SNP site were calculated. A linear regression was carried out for arcsine transformed allele frequencies as a function of latitude. Sites that were in complete linkage disequilibrium were considered as a single site for data analysis. Statistical analyses were carried out using R software (R Core Team 2013).

Results

Third chromosome extraction lines were generated for five populations (BMA, MPA, RNC, JFL and HFL). Luminosity measurements were made in female adult flies

for thoracic and abdominal pigmentation traits to determine whether the third chromosome contributes to variation in pigmentation within and among populations.

Phenotypic and geographic variation of pigments in extraction lines

Results of Nested ANOVA analysis revealed significant variation among the third chromosome extraction lines (F values ranged from 8.54 to 25.87, $P < 0.001$) and among the five locations sampled (F values ranged from 8.40 to 53.26, $P < 0.001$) (Table 3.1). A comparison of the phenotypic data from some of the extraction lines to the DGRP lines (n = 11) revealed a significant correlation in the thoracic trident (Table 3.2; correlation coefficient = .6862, $P < 0.05$). This implies that there is a detectable effect of factors on the third chromosome on thoracic trident pigmentation.

In order to test for clinal patterns in pigmentation, phenotype data from the extraction lines were regressed onto the latitude of origin for each population. Significant results were obtained for thoracic cuticle luminosity (t = - 2.038, $P < 0.05$) which was negatively associated with increases in latitude. Borderline significant results (t = -1,412, $P < 0.10$) were obtained with thoracic trident measurements (Table 3.3), which also showed a negative association. Negative but non-significant associations were uncovered for abdominal cuticle (t= -*.964, P > 0.10*) and the abdominal stripe (t= -*.270, P > 0.10*). A negative association between luminosity and latitude indicates that pigmentation increases with latitude (Figure 3.4).

Analysis of candidate sequences in extraction lines

Six sites in *pale* spanning 330 bp in a region 3057 bp upstream of the start codon were sequenced in extraction lines representative of populations in the eastern United States (BMA, MPA, RNC, JFL and HFL). These sites were previously associated with thoracic traits in the DGRP lines (See Chapter 2, Figure 2.5). Each site was polymorphic among the lines and four of the sites were completely linked (Table 3.4). The linked sites were considered as a single site in statistical analysis of SNP associations and clinal analysis. Each of the sites were significantly associated with thoracic traits ($8.26 < F < 33.80$, $P < 0.01, 0.001$). All of the sites were significantly associated with abdominal cuticle luminosity ($10.53 < F < 50.09$, $P < 0.01, 0.001$) while all but the four linked sites were significantly associated with abdominal stripe luminosity ($4.81 < F < 21.61$, $p < 0.05, 0.01, .001$) (Table 3.6). Allele frequencies for each site were calculated, arcsine transformed, and regressed unto latitude to determine if SNP sites exhibit clinal patterns. Results revealed a significant relationship between latitude and allele frequencies in site 3386 of *pale* ($t = 3.333$, $R^2 = .716$, $P < 0.05$) (Figure 3.5, Table 3.7). Allele frequencies in the rest of the sites were not significantly associated with latitude.

Nine sites in *ebony* spanning 542 bp in a region 3604 bp upstream of the start codon were also sequenced. Most of these sites were previously associated with abdominal traits while one was associated with thoracic traits in the DGRP lines (See Chapter 2, Figure 2.5). Each site was polymorphic among the lines and four of the sites were also completely linked (Table 3.5) to each other and thus were analyzed as a single site. Each of the sites had highly significant associations with both sets of thoracic traits ($21.83 < F < 199.20$, $p < 0.001$) and abdominal cuticle luminosity ($8.72 < F < 90.54$, $p <$

0.001). With regards to abdominal stripe luminosity, only two sites had a non-significant result (F = .28, 2.44) while the rest were significantly associated (*5.25* < F <*128.45, p <.05, p > 0.05, 0.01, 0.001*). Linear regression of allele frequencies of the candidate SNPS did not reveal any significant relationships with regards to latitude (Table 3.7).

Discussion

Phenotypic variation within natural populations

The major aim of this study was to determine whether DNA sequence variation in the candidate genes *pale* and *ebony* are associated with geographic patterns of pigmentation in North American populations of *D. melanogaster*. Phenotypic assays of third chromosome extraction lines representative of five populations in the eastern United States revealed significant variation in both thoracic and abdominal traits within and among these populations. Geographic location had a significant contribution to variance for all pigmentation traits in this study. These lines share a common genetic background except for their third chromosomes. Thus, the results of the assay demonstrate that this chromosome contributes significantly to phenotypic variation in pigmentation among populations of *D. melanogaster* in eastern United States.

Additionally, there appears to be a slight clinal pattern in pigmentation for each traits as luminosity is negatively (i.e. pigmentation is positively) associated with increases in latitude (Figure 3.4, Table 3.3). For thoracic traits, the effect of latitude was statistically significant, suggesting a stronger trend for this trait than the other traits. These findings reflect geographic patterns in *D. melanogaster* pigmentation that have also been uncovered in Asian (Munjal, Karan et al. 1997) and East Australian populations (Telonis-Scott, Hoffmann et al. 2011). Altitudinal but not clinal patterns were found in Sub-Saharan Africa, which is representative of the ancestral population of *Drosophila melanogaster*. These general patterns may reflect adaptation to similar

climatic conditions associated with higher altitudes and increasing latitudes in temperate environments. However, because *D. melanogaster* has been demonstrated to have undergone bottlenecks (Baudry, Viginier et al. 2004), patterns outside of African populations may reflect convergent evolution in pigmentation traits and genes in response to novel temperate climates. Desiccation tolerance, which is associated with pigmentation and exhibits clinal patterns in Asian populations (Parkash, Rajpurohit et al. 2008), may be a selective mechanism driving clinal variation in pigmentation. Darker individuals, which tended to be found in higher latitudes, were more tolerant to desiccation than lighter individuals. Clinal patterns in thermal sensitivity in *D. melanogaster* (Hoffmann, Anderson et al. 2002) as well as demonstrated roles of pigmentation in regulating body temperature in other insects (Watt 1968, Hirai and Kimura 1997) suggest that temperature gradients may also drive observed geographic patterns in pigmentation levels. Additionally, evidence for selection on pigmentation genes has been uncovered in *ebony*, where strong haplotype structure was associated with the upstream regulatory region of the gene in the darkest lines. This may be due to a recent selective sweep (Pool and Aquadro 2007).

In our study populations, thoracic pigmentation exhibited stronger evidence for clinal patterns than abdominal pigmentation. This suggests that thoracic traits may be more important than abdominal traits in responses to selective pressures that vary latitudinally in these populations. Increases in air temperature have been demonstrated to improve flight performance in insect species including *D. melanoagaster* (Lehmann 1999, Harrison and Roberts 2000). Therefore, the efficiency of warming thoracic indirect

flight muscles may impact flight performance and therefore the ability to find food and mates (Pringle 1949). As far as we know, this has yet to be directly tested. Abdominal traits do show significant non-clinal geographic variation, which may be due to distinct selective factors from thorax pigmentation. It is also possible that genes on the third chromosome simply have a greater influence on thoracic pigmentation than on abdominal pigmentation and that the distribution of causative alleles that influence thoracic pigmentation variation may be clinal as well. The significant correlation between the Raleigh DGRP and extraction lines in thoracic pigmentation is consistent with this conclusion.

Sequence variation among populations

Results of SNP associations in *ebony* and *pale* in the extraction lines were consistent with previous findings in the DGRP lines (See Chapter 2). Each candidate site was segregating multiple alleles which demonstrate that these sites are polymorphic throughout the tested populations and possibly across the range of *D. melanogaster* present in the eastern United States. In *pale*, the candidate sites sequenced in this study were previously associated with thoracic pigmentation traits (Figure 3.3) and were also associated with these traits in the extraction lines. Most of the sites were also associated with both sets of abdominal pigmentation traits. This provides further evidence that these SNPS may reside in a regulatory region that influences pigmentation intensity in these populations. One of these sites exhibited possible latitudinal variation and

exhibited strong associations with pigmentation traits, indicating that this site or linked sites may underlie geographic and latitudinal variation in pigmentation (Figure 3.5).

In *ebony*, most of the candidate sites were previously associated with abdominal traits with one site previously associated with thoracic traits (Figure 3.3). In the extraction lines there were strong associations of all sites with thoracic traits as well as abdominal traits. This region has already been identified as an enhancer for abdominal pigmentation in African populations (Rebeiz, Pool et al. 2009). The associations with thoracic traits found in this study might be due to the larger sample size and the increased power to detect effects and/or the controlled genetic background present in the extraction lines. The effects of these putative regulatory regions in *pale* and *ebony* in wild strains may be masked by the effects of other genes on the other chromosomes. My previous study with the DGRP lines detected significant SNP sites in *tan* and *Ddc* and evidence that these genes, along with *ebony* and *pale*, may all be involved in variation in pigmentation (See Chapter 2). None of the tested candidate SNPs in *ebony* presented evidence for latitudinal variation. Thus, these sites may only be contributing to local levels of variation in pigmentation traits.

The lack of clinal variation in these SNP sites may reflect limited sampling of sites and populations. However, while pigmentation assays do show some latitudinal trends, these are not very strong, which may reflect persistence of high levels of phenotypic variation in pigmentation traits within some of these populations as well as gene flow among populations. A closer look at the variance in these traits among the populations

may reveal that some populations have a higher propensity to maintain pigmentation variation. Fluctuations of temperature in higher latitudes, for instance, may cause temporal shifts in directions selection over time, thus acting to maintain pigmentation variation. Proposed roles of pigmentation in thermoregulation and desiccation tolerance suggest that climate may be a selective driver. Seasonal shifts in pigmentation have been demonstrated in *Drosophila jambulina* and were found to correlate with humidity levels and desiccation resistance (Parkash, Singh et al. 2009). Seasonal analysis of Indian populations of *D. melanogaster* in localities of high seasonal changes also demonstrated temporal differences in pigmentation, which was also found to have a strong genetic basis (Dev, Chahal et al. 2013).

Polymorphic analysis of the pigmentation genes in the DGRP lines provided evidence that balancing selection may be acting on non-coding regions and contributing to pigmentation diversity in these lines (see Chapter 2). Balancing selection may also be acting in the other populations in Eastern United States. It is also possible that factors other than pigmentation may contribute to adaptation to different latitudes and selection may be stronger on those traits. Another possibility is that there has not been enough time for selection to strongly diversify pigmentation traits and underlying genes among geographic populations. However, in *pale* at least, HKA results suggest that there is persistence of old polymorphisms in *D. melanogaster*. This demonstrates the propensity for polymorphisms to be maintained over time at least in the *pale* gene.

Overall, this study demonstrated geographic and possible latitudinal patterns in thoracic and abdominal pigmentation traits. Use of third chromosome extraction lines provided further evidence for the contribution of *pale* and *ebony* and possibly other loci on the third chromosome to variation in these traits. The candidate regions examined in both genes in this study appear to be possible regulatory regions that control levels gene expression in particular parts of the body. This supports the suggestion that pigmentation variation in *D. melanogaster* is associated with variation in non-coding regions of pigmentation genes. The association of SNPs appears to be stronger with thoracic traits, which may be more heavily influenced by genes on the third chromosome. Latitudinal variation of thoracic traits in extraction lines also supports this conclusion. The candidate site examined in *pale* may be driving this variation since at least one of the sites has a clinal pattern. However, more populations need to be tested in order to conclusively make this claim. Lastly, this work leads to the hypothesis that variation may be maintained by balancing selection on pigmentation genes, which in turn may be due to seasonal fluctuations and temporal shifts in selection.

References

Baudry, E., B. Viginier and M. Veuille (2004). "Non-African populations of *Drosophila melanogaster* have a unique origin." Molecular Biology and Evolution **21**(8): 1482-1491.

Blenau, W. and A. Baumann (2005). "Molecular characterization of the ebony gene from the American cockroach, *Periplaneta americana*." Archives of Insect Biochemistry and Physiology **59**(3): 184-195.

Brisson, J. A., D. C. D. Toni, I. Duncan and A. R. Templeton (2005). "Abdominal pigmentation variation in *Drosophila polymoprha*: Geographic variation in the trait, and underlying phylogeography." Evolution **59**(5): 1046-1059.

David, J. R., P. Capy, P. Véronique and S. Tsakas (1985). "Thoracic trident pigmentation in *Drosophila melanogaster*: Differentiation of geographical populations." Genetics Selection Evolution **17**(2): 211-224.

Dev, K., J. Chahal, R. Parkash and S. Kataria (2013). "Correlated changes in body melanisation and mating traits of *Drosophila melanogaster*: A seasonal analysis." Evolutionary Biology **40**(3): 366-376.

Futahashi, R., J. Sato, Y. Meng, S. Okamoto, T. Daimon, K. Yamamoto, Y. Suetsugu, J. Narukawa, H. Takahashi, Y. Banno, S. Katsuma, T. Shimada, K. Mita and H. Fujiwara (2008). "*yellow* and *ebony* are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*." Genetics **180**(4): 1995-2005.

Hargeby, A., J. Johansson and J. Ahnesjö (2004). "Habitat-specific pigmentation in a freshwater isopod: adaptive evolution over a small spatiotemporal scale." Evolution **58**(1): 81-94.

Harrison, J. F. and S. P. Roberts (2000). "Flight respiration and energetics." Annual Review of Physiology **62**(1): 179-205.

Hirai, Y. and M. T. Kimura (1997). "Incipient reproductive isolation between two morphs of *Drosophila elegans* (Diptera: Drosophilidae)." Biological Journal of the Linnean Society **61**(4): 501-513.

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

Hoffmann, A. A., A. Anderson and R. Hallas (2002). "Opposing clines for high and low temperature resistance in *Drosophila melanogaster*." Ecology Letters **5**(5): 614-618.

Huxley, J. (1976). "The coloration of *Papilio zalmoxis* and *P. antimachus*, and the discovery of tyndall blue in butterflies." Proceedings of the Royal Society of London. Series B. Biological Sciences **193**(1113): 441-453.

James, A. C., R. B. Azevedo and L. Partridge (1995). "Cellular basis and developmental timing in a size cline of *Drosophila melanogaster*." Genetics **140**(2): 659-666.

Kalmus, H. (1941). "The resistance to desiccation of *Drosophila* mutants affecting body colour." Proceedings of the Royal Society B: Biological Sciences **130**(859): 185-201.

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

Lehmann, F. O. (1999). "Ambient temperature affects free-flight performance in the fruit fly *Drosophila melanogaster*." Journal of Comparative Physiology B **169**(3): 165-171.

Miyazaki, S., Y. Okada, H. Miyakawa, G. Tokuda, R. Cornette, S. Koshikawa, K. Maekawa and T. Miura (2014). "Sexually dimorphic body color is regulated by sex-specific expression of *yellow* gene in Ponerine Ant, *Diacamma* Sp." PLoS ONE **9**(3): e92875.

Munjal, A. K., D. Karan, P. Gibert, B. Moreteau, R. Parkash and J. R. David (1997). "Thoracic trident pigmentation in *Drosophila melanogaster*: latitudinal and altitudinal clines in Indian populations." Genetics Selection Evolution **29**(6): 601-610.

Parkash, R., S. Rajpurohit and S. Ramniwas (2008). "Changes in body melanisation and desiccation resistance in highland vs. lowland populations of *Drosophila melanogaster*." J. Insect Physiol. Journal of Insect Physiology **54**(6): 1050-1056.

Parkash, R., S. Singh and S. Ramniwas (2009). "Seasonal changes in humidity level in the tropics impact body color polymorphism and desiccation resistance in *Drosophila jambulina*—Evidence for melanism-desiccation hypothesis." Journal of Insect Physiology **55**(4): 358-368.

Pool, J. E. and C. F. Aquadro (2007). "The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*." Molecular Ecology **16**(14): 2844-2851.

Pringle, J. (1949). "The excitation and contraction of the flight muscles of insects." The Journal of physiology **108**(2): 226-232.

R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Rajpurohit, S., R. Parkash and S. Ramniwas (2008). "Body melanization and its adaptive role in thermoregulation and tolerance against desiccating conditions in drosophilids." Entomological Research **38**(1): 49-60.

Rebeiz, M., J. E. Pool, V. A. Kassner, C. F. Aquadro and S. B. Carroll (2009). "Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population." Science **326**(5960): 1663-1667.

Schmidt, P. S. and A. B. Paaby (2008). "Reproductive diapause and life-history clines in North American populations of *Drosophila melanogaster*." Evolution **62**(5): 1204-1215.

Schmidt, P. S., C.-T. Zhu, J. Das, M. Batavia, L. Yang and W. F. Eanes (2008). "An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*." Proceedings of the National Academy of Sciences **105**(42): 16207-16211.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Meth **9**(7): 671-675.

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

Telonis-Scott, M., A. A. Hoffmann and C. M. Sgrò (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.

Watt, W. B. (1968). "Adaptive Significance of Pigment Polymorphisms in *Colias* Butterflies. I. Variation of Melanin Pigment in Relation to Thermoregulation." Evolution **22**(3): 437-458.

Watt, W. B. (1968). "Adaptive significance of pigment polymorphisms in Colias butterflies. I. Variation of melanin pigment in relation to thermoregulation." Evolution: 437-458.

Wittkopp, P. J. and P. Beldade (2009). "Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy." Seminars in Cell & Developmental Biology **20**(1): 65-71.

Wittkopp, P. J., E. E. Stewart, L. L. Arnold, A. H. Neidert, B. K. Haerum, E. M. Thompson, S. Akhras, G. Smith-Winberry and L. Shefner (2009). "Intraspecific Polymorphism to Interspecific Divergence: Genetics of Pigmentation in Drosophila." Science **326**(5952): 540-544.

Wright, T. R. (1987). "The genetics of biogenic amine metabolism, sclerotization, and melanization in Drosophila melanogaster." Adv. Genet **24**(2): 127.

Zhang, W., K. Kunte and M. R. Kronforst (2013). "Genome-Wide Characterization of Adaptation and Speciation in Tiger Swallowtail Butterflies Using De Novo Transcriptome Assemblies." Genome Biology and Evolution **5**(6): 1233-1245.

Tables and Figures

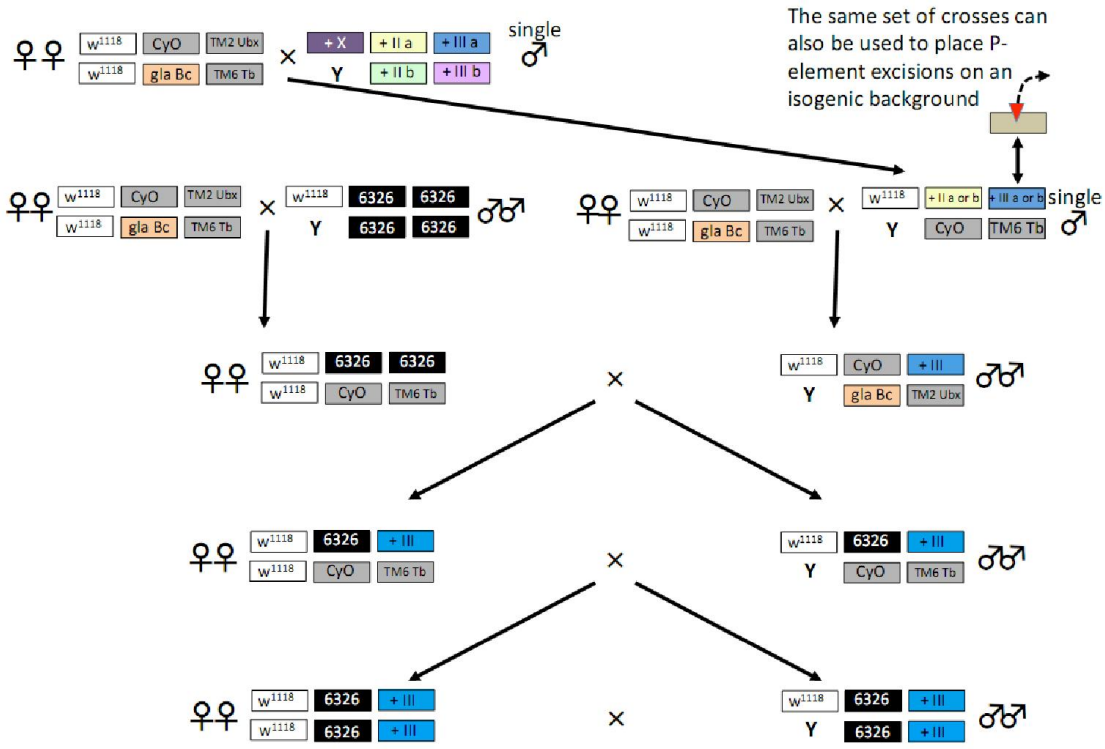


Figure 3.1. Schematic of crosses used to extract third chromosomes on isogenic backgrounds.

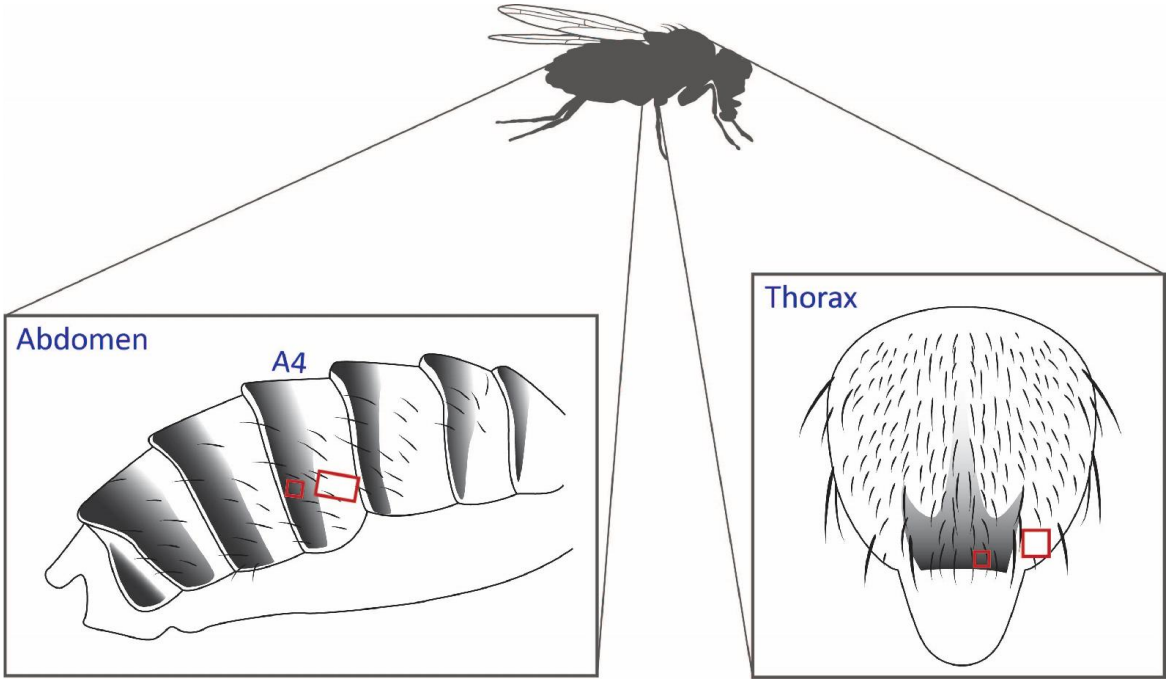


Figure 3.2. Diagram illustrating segments within thorax assayed for luminosity measurements. Red boxes indicates landmarks where measurements were taken in the A4 abdominal segment.

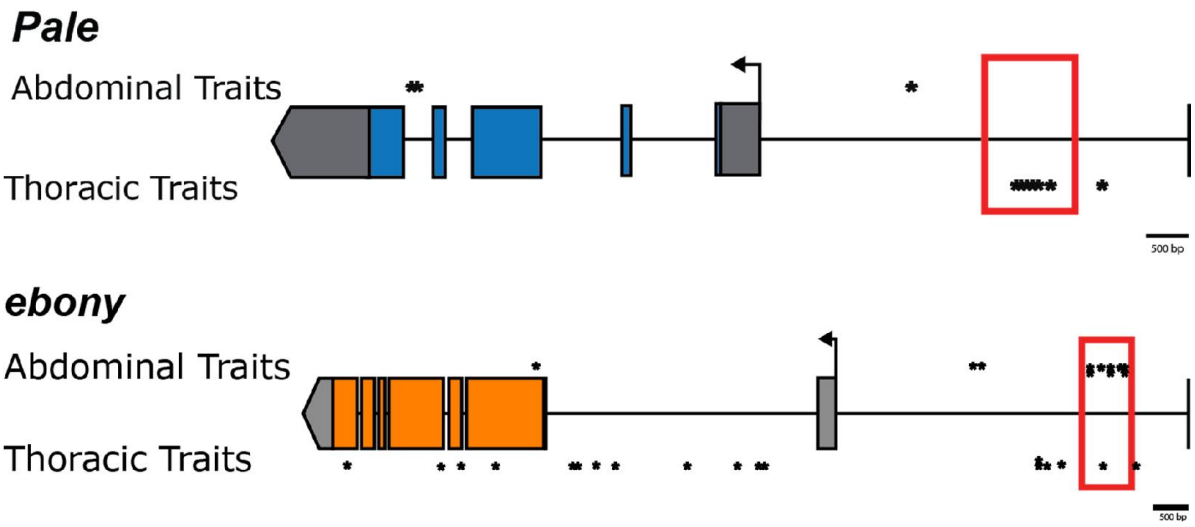


Figure 3.3. Gene models of *pale* and *ebony* with previously significant SNP sites associated with abdominal and thoracic pigmentation traits. Significant sites are denoted with asterisks. SNP sites sequenced and analyzed for this study are encased in squares outlined in red. Scales are provided in the bottom right corner of each model.

Table 3.1. Nested ANOVA Analysis of female pigmentation traits within third chromosome extraction lines representative of five populations. Lines were nested within locations. . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $p < 0.001$

Trait		df	SS	MS	*F*	*P* - value
Thoracic Cuticle	Location	4	0.0504	0.0126	44.58	**<.0001 *** **
	Line	93	0.3929	0.0042	14.95	**<.0001*** **
	Residuals	196	0.0554	0.0003		
Thoracic Trident	Location	4	0.0622	0.0155	47.52	**<.0001 *** **
	Line	93	0.7868	0.0085	25.87	**<.0001 *** **
	Residuals	196	0.0641	0.0003		
Abdominal Cuticle	Location	4	0.1271	0.0318	53.26	**<.0001 *** **
	Line	93	1.0683	0.0115	19.25	**<.0001 *** **
	Residuals	196	0.1170	0.0006		
Abdominal Stripe	Location	4	0.0077	0.0019	8.40	**<.0001 *** **
	Line	93	0.1814	0.0020	8.54	**<.0001 *** **
	Residuals	196	0.0448	0.0002		

Table 3.2. Result of Correlation Analysis of female pigmentation traits between GDRP lines and extraction lines. (n = 11). * $P < 0.05$, ** $P < 0.01$, *** $p < 0.001$

Trait	Correlation Coefficient	
Thoracic Cuticle	0.2964112	
Thoracic Trident	**0.6861797**	*
Abdominal Cuticle	0.3131454	
Abdominal Stripe	0.1705683	

Table 3.3. Results of a multi-level linear regression to determine if pigmentation cores vary predictably with latitude which would be indicative of a cline. . . . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $p < 0.001$

Fixed effects								
	Abdominal Cuticle				**Abdominal Stripe**			
Predictors	Coefficient	SE	*t*		Coefficient	SE	*t*	
Intercept	0.4702031	0.035057	**13.413**	*	0.104759	0.014005	7.480	
Latitude	-0.0009245	0.000959	-0.964		-0.0001036	0.000383	-0.270	

Variance Components (Random effects)				
	Abdominal Cuticle		**Abdominal Stripe**	
Parameter	Variance	SD	Variance	SD
Vial	0.0002672	0.01635	0.0001049	0.01024
Line	0.0039111	0.06254	0.0005798	0.02408

Fixed effects								
	Thoracic Cuticle				**Thoracic Trident**			
Predictors	Coefficient	SE	*t*		Coefficient	SE	*t*	
Intercept	0.2947302	0.021003	**14.033**	*	0.260159	0.029384	**8.854**	
Latitude	-0.0011713	0.000575	**-2.038**	*	-0.001135	0.000804	**-1.412**	.

Variance Components (Random effects)				
	Thoracic Cuticle		**Thoracic Trident**	
Parameter	Variance	SD	Variance	SD
Vial	9.13E-05	0.009557	0.0001377	0.01173
Line	1.38E-03	0.037164	0.0027789	0.05272

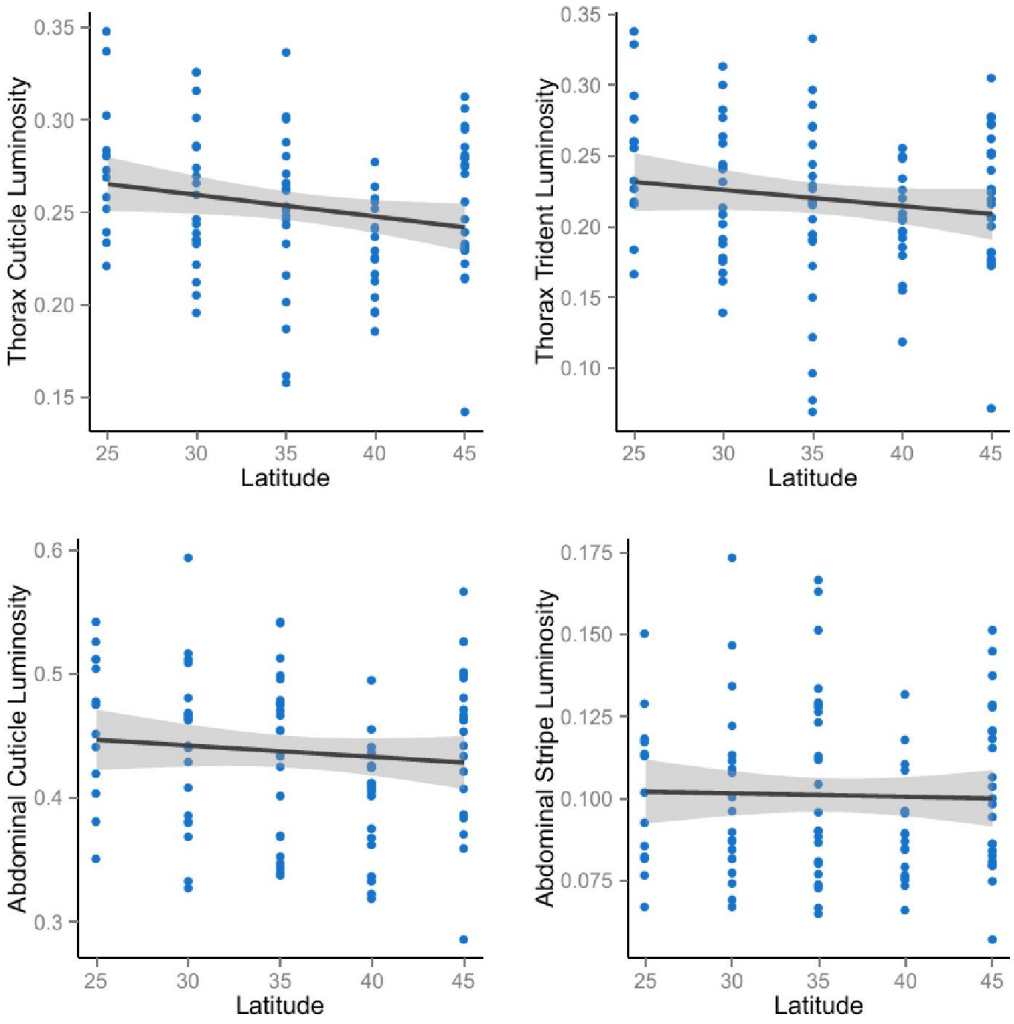


Figure 3.4. Luminosity of extraction lines plotted against latitude. Linear regression coefficient estimates were negative for each trait indicating decreasing luminosity or increasing pigmentation intensity with increases in latitude. Coefficients were statistically significant for thorax luminosity values ($P < 0.05$) and marginally significant for thorax trident values ($P < 0.10$)

Table 3.4. Candidate SNP sites in the *pale* gene in lines representative of the five tested populations in Eastern United States (Bowdoin, Maine, BME, Media, Pennsylvania, LPA, Raleigh, North Carolina, RNC Jacksonville, Florida, JFL and Homestead, Florida, HFL). SNP sites numbers refers to the bp position upstream of the start codon of *pale*. Dots denote matches in the sequence with the reference sequence obtained from Flybase, and letters denote the alternative nucleotide at each SNP site. The sites 3057, 3081, and 3100, and 3128 are in complete linkage.

			Candidate SNPS sites in *pale*					
Line	**Population**	**Latitude**	**3057**	**3081**	**3100**	**3128**	**3148**	**3386**
Reference	--	--	A	A	G	A	C	G
BME32	BME	45	.	.	.	.	.	.
BME9	BME	45	.	.	.	.	T	A
BME20	BME	45	.	.	.	.	T	A
BME33	BME	45	.	.	.	.	.	.
BME46	BME	45	.	.	.	.	T	A
BME47	BME	45	.	.	.	.	.	.
BME14	BME	45	.	.	.	.	T	A
BME22	BME	45	.	.	.	.	.	.
BME30	BME	45	.	.	.	.	T	A
BME15	BME	45	.	.	.	.	T	A
BME1	BME	45	T	T	A	G	.	.
BME42	BME	45	T	T	A	G	.	.
BME34	BME	45	T	T	A	G	.	.
BME19	BME	45	T	T	A	G	.	.
BME28	BME	45	T	T	A	G	.	.
BME23	BME	45	T	T	A	G	.	.
BME43	BME	45	T	T	A	G	.	.
BME48	BME	45	T	T	A	G	.	.
BME6	BME	45	T	T	A	G	.	.
LPA40	LPA	40	.	.	.	.	.	.
LPA20	LPA	40	.	.	.	.	.	A
LPA32	LPA	40	.	.	.	.	.	A
LPA51	LPA	40	.	.	.	.	.	.
LPA23	LPA	40	.	.	.	.	.	A
LPA12	LPA	40	.	.	.	.	T	A
LPA39	LPA	40	.	.	.	.	T	A
LPA34	LPA	40	.	.	.	.	.	.
LPA47	LPA	40	.	.	.	.	.	.
LPA24	LPA	40	.	.	.	.	.	.

LPA50	LPA	40	T	T	A	G	.	.
LPA48	LPA	40	T	T	A	G	.	.
LPA18	LPA	40	T	T	A	G	.	.
LPA30	LPA	40	T	T	A	G	.	.
LPA42	LPA	40	T	T	A	G	.	.
RNC334	RNC	35	.	.	.	.	.	A
RNC437	RNC	35	.	.	.	.	T	A
RNC712	RNC	35	.	.	.	.	T	A
RNC535	RNC	35	.	.	.	.	T	A
RNC774	RNC	35	.	.	.	.	T	A
RNC237	RNC	35	.	.	.	.	T	A
RNC59	RNC	35	T	T	A	G	.	.
RNC315	RNC	35	T	T	A	G	.	.
RNC865	RNC	35	T	T	A	G	.	.
RNC57	RNC	35	T	T	A	G	.	.
RNC109	RNC	35	T	T	A	G	.	.
RNC765	RNC	35	T	T	A	G	.	.
RNC507	RNC	35	T	T	A	G	.	.
JFL40	JFL	30	.	.	.	.	.	.
JFL39	JFL	30	.	.	.	.	.	A
JFL1	JFL	30	.	.	.	.	T	A
JFL7	JFL	30	.	.	.	.	.	.
JFL44	JFL	30	.	.	.	.	T	A
JFL46	JFL	30	.	.	.	.	T	A
JFL54	JFL	30	.	.	.	.	T	A
JFL49	JFL	30	.	.	.	.	.	A
JFL135	JFL	30	T	T	A	G	.	.
JFL120	JFL	30	T	T	A	G	.	.
JFL112	JFL	30	T	T	A	G	.	.
HFL117	HFL	25	.	.	.	.	T	A
HFL4	HFL	25	.	.	.	.	T	A
HFL134	HFL	25	.	.	.	.	.	.
HFL127	HFL	25	.	.	.	.	.	A
HFL19	HFL	25	.	.	.	.	T	A
HFL18	HFL	25	T	T	A	G	.	.
HFL6	HFL	25	T	T	A	G	.	.
HFL3	HFL	25	T	T	A	G	.	.

Table 3.5. Candidate SNP sites in the *ebony* gene in lines representative of the five tested populations in Eastern United States (Bowdoin, Maine, BME, Media, Pennsylvania, LPA, Raleigh, North Carolina, RNC Jacksonville, Florida, JFL and Homestead, Florida, HFL). SNP sites numbers refers to the bp position upstream of the start codon of *ebony*. Dots denote matches in the sequence with the reference sequence obtained from Flybase, and letters denote the alternative nucleotide at each SNP site. The sites 4058,4110,4134, and 4245 are in complete linkage. NAs denote ambiguous reads.

			Candidate SNP sites in *ebony*								
Line	**Pop**	**Latitude**	**3604**	**3777**	**3805**	**3828**	**3862**	**4058**	**4110**	**4134**	**4145**
Reference	--	--	T	T	G	T	A	G	T	T	A
BME42	BME	45	.	.	.	.	.	.	.	.	.
BME34	BME	45	A	.	T	.	.	A	A	A	T
BME20	BME	45	.	.	.	.	.	.	.	.	.
BME19	BME	45	.	.	.	.	.	.	.	.	.
BME32	BME	45	.	.	.	.	.	.	.	.	.
BME47	BME	45	.	.	.	.	.	.	.	.	.
BME23	BME	45	.	.	.	.	.	.	.	.	.
BME46	BME	45	.	.	.	.	.	.	.	.	.
BME9	BME	45	NA	.	.	.	.	.	.	.	.
BME33	BME	45	.	.	.	.	.	.	.	.	.
BME28	BME	45	.	.	.	.	.	.	.	.	.
BME14	BME	45	NA	.	.	.	.	.	.	.	.
BME48	BME	45	.	.	.	G	NA	.	.	.	NA
BME15	BME	45	.	.	.	G	NA	.	.	.	.
BME30	BME	45	.	.	.	.	.	.	.	.	.
BME6	BME	45	.	.	.	G	T	.	.	.	.
LPA47	LPA	40	A	.	T	.	.	A	A	A	T
LPA42	LPA	40	.	.	.	.	.	A	A	A	T
LPA50	LPA	40	.	.	.	.	.	.	.	.	.
LPA32	LPA	40	.	.	.	.	.	.	.	.	.
LPA51	LPA	40	.	.	.	.	.	.	.	.	.
LPA33	LPA	40	.	.	.	.	.	.	.	.	.
LPA20	LPA	40	.	.	.	G	T	.	.	.	.
LPA23	LPA	40	.	.	.	.	.	.	.	.	.
LPA39	LPA	40	.	.	.	G	NA	.	.	.	.
LPA12	LPA	40	.	.	.	.	.	.	.	.	.
LPA34	LPA	40	A	.	.	.	.	.	.	.	.
LPA30	LPA	40	.	G	.	.	.	.	.	.	.
LPA18	LPA	40	.	.	.	G	T	.	.	.	.
LPA24	LPA	40	.	.	.	.	.	.	.	.	.

LPA38	LPA	40	.	.	.	G	T	.	.	.	.
RNC380	RNC	35	A	.	T	.	.	A	A	A	T
RNC507	RNC	35	A	.	T	.	.	A	A	A	T
RNC237	RNC	35	A	.	T	.	.	A	A	A	T
RNC59	RNC	35	.	G	.	.	.	.	.	.	.
RNC334	RNC	35	.	G	.	.	.	.	.	.	.
RNC437	RNC	35	.	G	.	.	.	.	.	.	.
RNC712	RNC	35	.	G	.	.	.	.	.	.	.
RNC315	RNC	35	.	.	.	.	.	.	.	.	.
RNC109	RNC	35	.	.	.	.	.	.	.	.	.
RNC320	RNC	35	.	.	.	.	.	.	.	.	.
RNC379	RNC	35	.	.	.	.	.	.	.	.	.
RNC365	RNC	35	.	G	.	.	.	.	.	.	.
RNC865	RNC	35	.	G	.	.	.	.	.	.	.
RNC765	RNC	35	.	.	.	.	.	.	.	.	.
RNC313	RNC	35	.	.	.	.	.	.	.	.	.
RNC535	RNC	35	.	.	.	.	.	.	.	.	.
RNC774	RNC	35	.	.	.	G	NA	.	.	.	NA
JFL35	JFL	30	.	.	.	.	.	.	.	.	.
JFL39	JFL	30	.	.	.	.	.	.	.	.	.
JFL40	JFL	30	.	.	.	.	.	.	.	.	.
JFL1	JFL	30	.	.	.	.	.	.	.	.	.
JFL135	JFL	30	NA	.	.	.	.	.	.	.	.
JFL38	JFL	30	.	.	.	.	.	.	.	.	.
JFL112	JFL	30	.	.	.	.	.	.	.	.	.
JFL22	JFL	30	.	.	.	.	.	.	.	.	.
JFL7	JFL	30	A	.	.	.	.	.	.	.	.
JFL49	JFL	30	.	G	.	.	.	.	.	.	.
JFL46	JFL	30	.	G	.	.	.	.	.	.	.
JFL54	JFL	30	A	.	.	.	.	.	.	.	.
JFL51	JFL	30	.	.	.	G	T	.	.	.	.
HFL117	HFL	25	.	.	.	G	NA	.	.	.	.
HFL18	HFL	25	.	.	.	.	.	.	.	.	.
HFL127	HFL	25	.	.	.	G	NA	.	.	.	NA

Table 3.6. Results of SNP association analysis of sequenced candidate SNPS in *ebony* and *pale* among all lines representative of tested populations (Bowdoin, Maine, Media, Pennsylvania, Raleigh, North Carolina, Jacksonville, Florida, and Homestead, Florida). SNP sites numbers refers to the bp position upstream of the start codon of *pale* or *ebony*. The sites 3057, 3081, and 3100, and 3128 are in complete linkage in *pale*. The sites 4058,4110,4134, and 4245 in *ebony* are in complete linkage in *ebony*. These were analyzed as single sites in this analysis. F-values from ANOVA tests are reported for each site and associated traits. . $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $p < 0.001$

Gene	Sites	Thoracic Cuticle		Thoracic Trident		Abdominal Cuticle		Abdominal Stripe	
Pale	3057-3128	**8.26**	**	**9.30**	**	**5.46**	*	0.15	
	3148	**30.05**	***	**33.80**	***	**21.61**	***	**4.81**	*
	3386	**15.81**	***	**13.36**	***	**8.75**	**	**6.42**	*
ebony	3604	**49.51**	***	**114.34**	***	**12.70**	***	**22.17**	***
	3777	**28.95**	***	**86.87**	***	**20.10**	***	0.28	
	3805	**21.83**	***	**45.46**	***	**15.79**	***	2.44	
	3828	**113.78**	***	**199.20**	***	**103.38**	***	**69.00**	***
	3862	**78.35**	***	**184.10**	***	**128.45**	***	**36.93**	***
	4058-4245	**44.86**	***	**68.52**	***	**6.09**	*	**5.25**	*

Table 3.7. Results of linear regressions of latitude and allele frequencies at candidate SNP sites in *ebony* and *pale* in among lines representative of tested populations (Bowdoin, Maine, Media, Pennsylvania, Raleigh, North Carolina, Jacksonville, Florida, and Homestead, Florida). SNP sites numbers refers to the bp position upstream of the start codon of *pale* or *ebony*. The sites 3057, 3081, and 3100, and 3128 are in complete linkage in *pale*. The sites 4058,4110,4134, and 4245 in *ebony* are in complete linkage in *ebony*. These were analyzed as single sites in this analysis. Allele frequencies were arcsine transformed for this analysis.
. $P < 0.10$, * $P < 0.05$, ** $P < 0.01$, *** $p < 0.001$

Gene	SNP sites	*Estimate*	*SE*	*t*	R^2	*P*	
Pale	3057-3128	-16.96	38.83	-0.437	-0.254	0.692	
	3148	34.13	41.86	0.815	-0.091	0.475	
	3386	**78.81**	**23.65**	**3.333**	**0.716**	**0.045**	*
ebony	3604	-18.74	23.38	-0.801	-0.098	0.481	
	3777	-7.94	18.98	-0.419	-0.260	0.704	
	3805	-27.11	18.90	-1.434	0.209	0.247	
	3828	12.08	19.97	0.605	-0.189	0.588	
	3862	-23.14	19.14	-1.209	0.104	0.313	
	4058-4245	-26.80	16.81	-1.594	0.278	0.209	

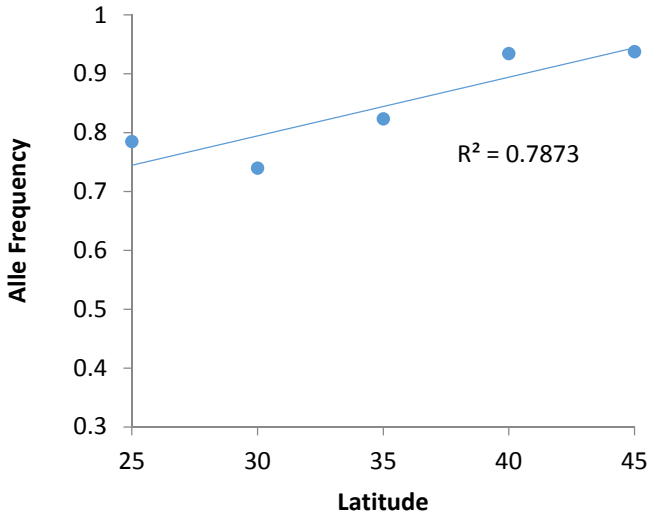


Figure 3.5. Regression of allele frequencies in SNP site 3386 in *pale* against latitude. Linear regression results for this site were significant ($P < 0.05$, $R^2 = 0.716$) Allele frequencies were arcsine transformed for this analysis.

References

Alekseyenko, O. V., Y.-B. Chan, R. Li and E. A. Kravitz (2013). "Single dopaminergic neurons that modulate aggression in *Drosophila*." Proceedings of the National Academy of Sciences **110**(15): 6151-6156.

Andersen, S. O. (2010). "Insect cuticular sclerotization: A review." Insect Biochemistry and Molecular Biology **40**(3): 166-178.

Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollmann, L. H. Duncan, F. Lawrence, R. R. H. Anholt and T. F. C. Mackay (2009). "Systems genetics of complex traits in *Drosophila melanogaster*." Nat Genet **41**(3): 299-307.

Barnes, A. I. and M. T. Siva-Jothy (2000). "Density-dependent prophylaxis in the mealworm beetle *Tenebrio molitor* L. (Coleoptera: Tenebrionidae): cuticular melanization is an indicator of investment in immunity." Proc Biol Sci **267**(1439): 177-182.

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

Baudry, E., B. Viginier and M. Veuille (2004). "Non-African populations of *Drosophila melanogaster* have a unique origin." Molecular Biology and Evolution **21**(8): 1482-1491.

Benovoy, D., T. Kwan and J. Majewski (2008). "Effect of polymorphisms within probe–target sequences on olignonucleotide microarray experiments." Nucleic Acids Research **36**(13): 4417-4423.

Biessmann, H. (1985). "Molecular analysis of the *yellow* gene (*y*) region of *Drosophila melanogaster*." Proceedings of the National Academy of Sciences **82**(21): 7369-7373.

Birman, S., B. Morgan, M. Anzivino and J. Hirsh (1994). "A novel and major isoform of tyrosine hydroxylase in *Drosophila* is generated by alternative RNA processing." Journal of Biological Chemistry **269**(42): 26559-26567.

Blenau, W. and A. Baumann (2005). "Molecular characterization of the ebony gene from the American cockroach, *Periplaneta americana*." Archives of Insect Biochemistry and Physiology **59**(3): 184-195.

Brisson, J. A., D. C. D. Toni, I. Duncan and A. R. Templeton (2005). "Abdominal pigmentation variation in *Drosophila polymoprha*: Geographic variation in the trait, and underlying phylogeography." Evolution **59**(5): 1046-1059.

Brunetti, C. R., J. E. Selegue, A. Monteiro, V. French, P. M. Brakefield and S. B. Carroll (2001). "The generation and diversification of butterfly eyespot color patterns." Current Biology **11**(20): 1578-1585.

Chen, L., G. P. Page, T. Mehta, R. Feng and X. Cui (2009). "Single nucleotide polymorphisms affect both cis-and trans-eQTLs." Genomics **93**(6): 501-508.

Cloudsley-Thompson, J. L. (1999). "Multiple factors in the evolution of animal coloration." Naturwissenschaften **86**(3): 123-132.

Cochran, W. G. (1954). "Some methods for strengthening the common χ *2* tests." Biometrics **10**(4): 417-451.

Consortium, D. G. (2007). "Evolution of genes and genomes on the *Drosophila* phylogeny." Nature **450**(7167): 203-218.

Cridland, J. M., S. J. Macdonald, A. D. Long and K. R. Thornton (2013). "Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources." Molecular Biology and Evolution **30**(10): 2311-2327.

David, J. R. and P. Capy (1988). "Genetic variation of *Drosophila melanogaster* natural populations." Trends in Genetics **4**(4): 106-111.

David, J. R., P. Capy, P. Véronique and S. Tsakas (1985). "Thoracic trident pigmentation in *Drosophila melanogaster*: Differentiation of geographical populations." Genetics Selection Evolution **17**(2): 211-224.

Dev, K., J. Chahal, R. Parkash and S. Kataria (2013). "Correlated changes in body melanisation and mating traits of *Drosophila melanogaster*: A seasonal analysis." Evolutionary Biology **40**(3): 366-376.

Dombeck, I. and J. Jaenike (2004). "Ecological genetics of abdominal pigmentation in *Drosophila falleni*: A pleiotropic link to nematode parasitism." Evolution **58**(3): 587-596.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.
Falconer, D. S. and T. F. C. Mackay (1996). Introduction to quantitative genetics. Essex, England, Longman.

Fay, J. C. and C.-I. Wu (2000). "Hitchhiking under positive Darwinian selection." Genetics **155**(3): 1405-1413.

Flowers, J., E. Sezgin, S. Kumagai, D. Duvernell, L. Matzkin, P. Schmidt and W. Eanes (2007). "Adaptive evolution of metabolic pathways in *Drosophila*." Molecular Biology and Evolution **24**(6): 1347-1354.

Futahashi, R., J. Sato, Y. Meng, S. Okamoto, T. Daimon, K. Yamamoto, Y. Suetsugu, J. Narukawa, H. Takahashi, Y. Banno, S. Katsuma, T. Shimada, K. Mita and H. Fujiwara (2008). "*yellow* and *ebony* are the responsible genes for the larval color mutants of the silkworm *Bombyx mori*." Genetics **180**(4): 1995-2005.

Gompel, N. and S. B. Carroll (2003). "Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies." Nature **424**(6951): 931-935.

Hargeby, A., J. Johansson and J. Ahnesjö (2004). "Habitat-specific pigmentation in a freshwater isopod: adaptive evolution over a small spatiotemporal scale." Evolution **58**(1): 81-94.

Harrison, J. F. and S. P. Roberts (2000). "Flight respiration and energetics." Annual Review of Physiology **62**(1): 179-205.

Hegmann, J. P. and B. Possidente (1981). "Estimating genetic correlations from inbred strains." Behavior genetics **11**(2): 103-114.

Hey, J. (2001). HKA Program. Rutgers University.

Hirai, Y. and M. T. Kimura (1997). "Incipient reproductive isolation between two morphs of *Drosophila elegans* (Diptera: Drosophilidae)." Biological Journal of the Linnean Society **61**(4): 501-513.

Hirsh, J., B. A. Morgan and S. B. Scholnick (1986). "Delimiting regulatory sequences of the *Drosophila melanogaster Ddc* gene." Molecular and Cellular Biology **6**(12): 4548-4557.

Hirsh, J., T. Riemensperger, H. Coulom, M. Iché, J. Coupar and S. Birman (2010). "Roles of dopamine in circadian rhythmicity and extreme light sensitivity of circadian entrainment." Current Biology **20**(3): 209-214.

Hoekstra, H. E., J. G. Krenz and M. W. Nachman (2004). "Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations." Heredity **94**(2): 217-228.

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

Hoffmann, A. A., A. Anderson and R. Hallas (2002). "Opposing clines for high and low temperature resistance in *Drosophila melanogaster*." Ecology Letters **5**(5): 614-618.

Hovemann, B. T., R. P. Ryseck, U. Walldorf, K. F. Stortkuhl, I. D. Dietzel and E. Dessen (1998). "The *Drosophila ebony* gene is closely related to microbial peptide synthetases and shows specific cuticle and nervous system expression." Gene **221**(1): 1-9.

Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Rámia, A. Tarone, L. Turlapati, T. Zichner, D. Zhu, R. Lyman, M. Magwire, K. Blankenburg, M. A. Carbone, K. Chang, L. Ellis, S. Fernandez, Y. Han, G. Highnam, C. Hjelmen, J. Jack, M. Javaid, J. Jayaseelan, D. Kalra, S. Lee, L. Lewis, M. Munidasa, F. Ongeri, S. Patel, L. Perales, A. Perez, L. Pu, S. Rollmann, R. Ruth, N. Saada, C. Warner, A. Williams, Y.-Q. Wu, A. Yamamoto, Y. Zhang, Y. Zhu, R. Anholt, J. Korbel, D. Mittelman, D. Muzny, R. Gibbs, A. Barbadilla, S. Johnston, E. Stone, S. Richards, B. Deplancke and T. Mackay (2014). "Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines." Genome Research.

Hudson, R. R., M. Kreitman and M. Aguadé (1987). "A test of neutral molecular evolution based on nucleotide data." Genetics **116**(1): 153-159.

Huxley, J. (1976). "The coloration of *Papilio zalmoxis* and *P. antimachus*, and the discovery of tyndall blue in butterflies." Proceedings of the Royal Society of London. Series B. Biological Sciences **193**(1113): 441-453.

James, A. C., R. B. Azevedo and L. Partridge (1995). "Cellular basis and developmental timing in a size cline of *Drosophila melanogaster*." Genetics **140**(2): 659-666.

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

Jeong, S., M. Rebeiz, P. Andolfatto, T. Werner, J. True and S. B. Carroll (2008). "The Evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species." Cell **132**(5): 783-793.

Joron, M., C. D. Jiggins, A. Papanicolaou and W. O. McMillan (2006). "*Heliconius* wing patterns: an evo-devo model for understanding phenotypic diversity." Heredity **97**(3): 157-167.

Kalmus, H. (1941). "The resistance to desiccation of *Drosophila* mutants affecting body colour." Proceedings of the Royal Society of London Series B-Biological Sciences **130**(859): 185-201.

Kopp, A. and J. R. True (2002). "Evolution of male sexual characters in the Oriental *Drosophila melanogaster* species group." Evolution & Development **4**(4): 278-291.

Kopp, A., I. Duncan and S. B. Carroll (2000). "Genetic control and evolution of sexually dimorphic characters in *Drosophila*." Nature **408**(6812): 553-559.

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

Kyriacou, C. P., B. Burnet and K. Connolly (1978). "The behavioural basis of overdominance in competitive mating success at the ebony locus of *Drosophila melanogaster*." Animal Behaviour **26**(Part 4): 1195-1206.

Lavington, E., R. Cogni, C. Kuczynski, S. Koury, E. L. Behrman, K. R. O'Brien, P. S. Schmidt and W. F. Eanes (2014). "A small system—high-resolution study of metabolic adaptation in the central metabolic pathway to temperate climates in *Drosophila melanogaster*." Molecular Biology and Evolution.

Lehmann, F. O. (1999). "Ambient temperature affects free-flight performance in the fruit fly *Drosophila melanogaster*." Journal of Comparative Physiology B **169**(3): 165-171.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics **25**(11): 1451-1452.

Linheiro, R. S. and C. M. Bergman (2012). "Whole genome resequencing reveals natural target site preferences of transposable elements in *Drosophila melanogaster*." PLoS One **7**(2): e30008.

Loewe, L. and B. Charlesworth (2006). "Inferring the distribution of mutational effects on fitness in *Drosophila*." Biology letters **2**(3): 426-430.

Loewe, L., B. Charlesworth, C. Bartolomé and V. Nöel (2006). "Estimating Selection on Nonsynonymous Mutations." Genetics **172**(2): 1079-1092.

Long, A. D., R. F. Lyman, C. H. Langley and T. F. C. Mackay (1998). "Two Sites in the delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster*." Genetics **149**(2): 999-1017.

Lu, Y. and M. D. Rausher (2003). "Evolutionary Rate Variation in Anthocyanin Pathway Genes." Molecular Biology and Evolution **20**(11): 1844-1853.
Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barron, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L.-L. Pu, C. Qu, M. Ramia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley, Y.-Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman and R. A. Gibbs (2012). "The *Drosophila melanogaster* Genetic Reference Panel." Nature **482**(7384): 173-178.

Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire and J. M. Cridland (2012). "The *Drosophila melanogaster* genetic reference panel." Nature **482**(7384): 173-178.

Majerus, M. E. N. (1998). Melanism : evolution in action. Oxford ; New York, Oxford University Press.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet and C. D. Jiggins (2013). "Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies." Genome Research **23**(11): 1817-1828.

Marygold, S. J., P. C. Leyland, R. L. Seal, J. L. Goodman, J. Thurmond, V. B. Strelets and R. J. Wilson (2013). "FlyBase: improvements to the bibliography." Nucleic Acids Res **41**(Database issue): D751-757.

Massouras, A., S. M. Waszak, M. Albarca-Aguilera, K. Hens, W. Holcombe, J. F. Ayroles, E. T. Dermitzakis, E. A. Stone, J. D. Jensen, T. F. C. Mackay and B. Deplancke (2012). "Genomic variation and its impact on gene expression in *Drosophila melanogaster*." PLoS Genet **8**(11): e1003055.

McDonald, J. H. and M. Kreitman (1991). "Adaptive protein evolution at the *Adh* locus in *Drosophila*." Nature **351**(6328): 652-654.

Miyazaki, S., Y. Okada, H. Miyakawa, G. Tokuda, R. Cornette, S. Koshikawa, K. Maekawa and T. Miura (2014). "Sexually dimorphic body color is regulated by sex-specific expression of *yellow* gene in Ponerine Ant, *Diacamma* Sp." PLoS ONE **9**(3): e92875.

Morgan, T. H. and C. B. Bridges (1916). Sex-linked inheritance in *Drosophila*. Carnegie Institution of Washington publication 237. Washington, Carnegie Institution of Washington,.

Munjal, A. K., D. Karan, P. Gibert, B. Moreteau, R. Parkash and J. R. David (1997). "Thoracic trident pigmentation in *Drosophila melanogaster*: latitudinal and altitudinal clines in Indian populations." Genetics Selection Evolution **29**(6): 601-610.

Neckameyer, W. S. and K. White (1993). "*Drosophila* Tyrosine Hydroxylase is Encoded by the *pale* Locus." Journal of Neurogenetics **8**(4): 189-199.

Olson-Manning, C. F., C.-R. Lee, M. D. Rausher and T. Mitchell-Olds (2013). "Evolution of flux control in the glucosinolate pathway in *Arabidopsis thaliana*." Molecular Biology and Evolution **30**(1): 14-23.

Parkash, R., B. Kalra and V. Sharma (2010). "Impact of body melanisation on contrasting levels of desiccation resistance in a circumtropical and a generalist *Drosophila* species." Evolutionary Ecology **24**(1): 207-225.

Parkash, R., S. Rajpurohit and S. Ramniwas (2008). "Changes in body melanisation and desiccation resistance in highland vs. lowland populations of *Drosophila melanogaster*." J. Insect Physiol. Journal of Insect Physiology **54**(6): 1050-1056.

Parkash, R., S. Singh and S. Ramniwas (2009). "Seasonal changes in humidity level in the tropics impact body color polymorphism and desiccation resistance in *Drosophila jambulina*—Evidence for melanism-desiccation hypothesis." Journal of Insect Physiology **55**(4): 358-368.

Pendleton, R., A. Rasheed, T. Sardina, T. Tully and R. Hillman (2002). "Effects of Tyrosine Hydroxylase mutants on locomotor activity in *Drosophila*: A study in functional genomics." Behavior Genetics **32**(2): 89-94.

Pérez, M. M., J. Schachter, J. Berni and L. A. Quesada-Allué (2010). "The enzyme NBAD-synthase plays diverse roles during the life cycle of *Drosophila* melanogaster." Journal of Insect Physiology **56**(1): 8-13.

Pool, J. E. and C. F. Aquadro (2007). "The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*." Molecular Ecology **16**(14): 2844-2851.

Pringle, J. (1949). "The excitation and contraction of the flight muscles of insects." The Journal of physiology **108**(2): 226-232.

Protas, M. E. and N. H. Patel (2008). "Evolution of Coloration Patterns." Annual Review of Cell and Developmental Biology **24**: 425-446.

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

R Core Team (2013). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Rajpurohit, S., R. Parkash and S. Ramniwas (2008). "Body melanization and its adaptive role in thermoregulation and tolerance against desiccating conditions in drosophilids." Entomological Research **38**(1): 49-60.

Rebeiz, M., J. E. Pool, V. A. Kassner, C. F. Aquadro and S. B. Carroll (2009). "Stepwise modification of a modular enhancer underlies adaptation in a *Drosophila* population." Science **326**(5960): 1663-1667.

Riemensperger, T., G. Isabel, H. Coulom, K. Neuser, L. Seugnet, K. Kume, M. Iché-Torres, M. Cassar, R. Strauss, T. Preat, J. Hirsh and S. Birman (2011). "Behavioral consequences of dopamine deficiency in the *Drosophila* central nervous system." Proceedings of the National Academy of Sciences **108**(2): 834-839.

Rosenblum, E. B., H. E. Hoekstra and M. W. Nachman (2004). "Adaptive reptile color variation and the evolution of the *Mc1R* gene." Evolution **58**(8): 1794-1808.

Schmidt, P. S. and A. B. Paaby (2008). "Reproductive diapause and life-history clines in North American populations of *Drosophila melanogaster*." Evolution **62**(5): 1204-1215.

Schmidt, P. S., C.-T. Zhu, J. Das, M. Batavia, L. Yang and W. F. Eanes (2008). "An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*." Proceedings of the National Academy of Sciences **105**(42): 16207-16211.

Schmidt, P. S., L. Matzkin, M. Ippolito and W. F. Eanes (2005). "Geographic variation in diapause incident, life-history traits and climatic adaptation in *Drosophila melanogaster*." Evolution **59**(8): 1721-1732.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Meth **9**(7): 671-675.

Tajima, F. (1983). "Evolutionary relationship of DNA sequences in finite populations." Genetics **105**(2): 437-460.

Tajima, F. (1989). "Statistical method for testing the neutral mutation hypothesis by DNA polymorphism." Genetics **123**(3): 585-595.

Takahashi, A. Y. A. and T. Takano-Shimizu (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.

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

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0." Molecular Biology and Evolution.

Telonis-Scott, M., A. A. Hoffmann and C. M. Sgrò (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.

True, J. R. (2003). "Insect melanism: the molecules matter." Trends in Ecology & Evolution **18**(12): 640-647.

Vavricka, C., B. Christensen and J. Li (2010). "Melanization in living organisms: a perspective of species evolution." Protein & Cell **1**(9): 830-841.

Watt, W. B. (1968). "Adaptive significance of pigment polymorphisms in *Colias* butterflies. I. Variation of melanin pigment in relation to thermoregulation." Evolution **22**(3): 437-458.

Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." Theoretical Population Biology **7**(2): 256-276.

Wiernasz, D. C. (1989). "Female choice and sexual selection of male wing melanin pattern in *Pieris occidentalis* (Lepidoptera)." Evolution **43**(8): 1672-1682.

Wittkopp, P. J. and P. Beldade (2009). "Development and evolution of insect pigmentation: Genetic mechanisms and the potential consequences of pleiotropy." Seminars in Cell & Developmental Biology **20**(1): 65-71.

Wittkopp, P. J., B. L. Williams, J. E. Selegue and S. B. Carroll (2003). "*Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes." Proceedings of the National Academy of Sciences of the United States of America **100**(4): 1808-1813.

Wittkopp, P. J., E. E. Stewart, L. L. Arnold, A. H. Neidert, B. K. Haerum, E. M. Thompson, S. Akhras, G. Smith-Winberry and L. Shefner (2009). "Intraspecific

polymorphism to interspecific divergence: genetics of pigmentation in *Drosophila*." Science **326**(5952): 540-544.

Wittkopp, P. J., J. R. True and S. B. Carroll (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., K. Vaccaro and S. B. Carroll (2002). "Evolution of *yellow* gene regulation and pigmentation in *Drosophila*." Current Biology **12**(18): 1547-1556.

Wittkopp, P. J., S. B. Carroll and A. Kopp (2003). "Evolution in black and white: genetic control of pigment patterns in *Drosophila*." Trends in Genetics **19**(9): 495-504.

Wright, K. M. and M. D. Rausher (2010). "The evolution of control and distribution of adaptive mutations in a metabolic pathway." Genetics **184**(2): 483-502.

Yeh, S. D., S. R. Liou and J. R. True (2006). "Genetics of divergence in male wing pigmentation and courtship behavior between *Drosophila elegans* and *Drosophila gunungcola*." Heredity **96**(5): 383-395.

Yu, H.-S., Y.-H. Shen, G.-X. Yuan, Y.-G. Hu, H.-E. Xu, Z.-H. Xiang and Z. Zhang (2011). "Evidence of selection at melanin synthesis pathway loci during silkworm domestication." Molecular Biology and Evolution **28**(6): 1785-1799.

Zeng, K., Y.-X. Fu, S. Shi and C.-I. Wu (2006). "Statistical Tests for Detecting Positive Selection by Utilizing High-Frequency Variants." Genetics **174**(3): 1431-1439.

Zhang, W., K. Kunte and M. R. Kronforst (2013). "Genome-wide characterization of adaptation and speciation in tiger swallowtail butterflies using de novo transcriptome assemblies." Genome Biology and Evolution **5**(6): 1233-1245.