DEMBECK, LAUREN MARIE. Quantitative Genetics of Cuticular Traits in *Drosophila melanogaster*. (Under the direction of Dr. Trudy F. C. Mackay).

The cuticle serves as a boundary between the internal and external environments of an insect. The hardened melanin and sclerotin that comprise the exocuticle and color the body provide a protective shield and aid in thermoregulation and resistance to ultra-violet radiation from the sun. On the epicuticle is a layer of fatty-acid derived lipids. The ability to synthesize this waxy barrier that prevents desiccation was an evolutionary innovation and allowed insects to colonize land. In *Drosophila melanogaster* these lipids are primarily cuticular hydrocarbons (CHCs). CHCs not only provide water-proofing. Some of them have been co-opted in *D. melanogaster* and other insects as contact pheromones. In order to understand how these traits evolve, we must first identify the genetic variants that underlie natural variation within populations. I used the recently developed *Drosophila melanogaster* Genetic Reference Panel (DGRP) to identify genetic variants contributing to natural variation in pigmentation traits and CHC composition.

For Chapter 2, I conducted a genome-wide association (GWA) study on the proportion of dark melanin present on the two posterior abdominal tergites, tergites 5 (T5) and 6 (T6), in *D. melanogaster* females. I characterized significant genetic variation in the proportion of melanization and found high broad-sense heritability for each tergite. GWA analyses identified over 150 DNA variants associated with the proportion of melanization on T5, T6, and the difference between T5 and T6. Several variants were located within cis-regulatory regions of three genes known to contribute to variation in female body pigmentation, *bric-a-brac1*, *tan*, and *ebony*. I also tested 28 novel candidate genes for effects on these pigmentation traits with RNA interference (RNAi) and mutant analyses. Seventeen of 28 genes affected body coloration. Several of these genes are involved in developmental and regulatory pathways, chitin production, cuticle structure, and vesicle formation and transport. These findings showed that genetic variation may affect multiple steps in pathways involved in tergite development and melanization.

For Chapter 3, I conducted a second pigmentation study focused on the intensity of the color of sclerotin on tergite 3 and of melanin on tergite 6. I dissected, mounted, and phenotyped the abdominal cuticles from DGRP females and conducted GWA analyses. There were 72 genetic variants in or near 58 genes. We demonstrated that four of these genes affected body coloration in mutant and RNAi knockdown analyses. One of these was *Megalin*, a gene known to regulate the biosynthesis of black melanin in the *D. melanogaster* wing by endocytic removal of Yellow. Our results suggest these genes may be involved in the epidermis-cuticle vesicle dynamics that paint the *D. melanogaster* body.

For chapter 4, I collected and analyzed gas chromatography spectra from DGRP females and males and quantified relative abundance of CHCs. Clusters of CHCs are highly correlated; therefore,
I conducted GWA analyses on the CHC principal components (PC). There were 305 and 173 unique genes containing or near significant genetic variants in the female and male PC GWA analyses, respectively. I selected 24 candidate genes for functional analysis using RNAi knockdown and one mutant line. These genes included cytochrome P450s, peroxidases, a palmitoyltransferase, thiolester and carboxylic ester hydrolases, fatty acid elongases, fatty acyl-CoA reductases, and a NADH dehydrogenase. All genes tested had significant effects on CHC composition. Two knockdown lines dramatically increased the total amounts of CHCs in both sexes. I also found that the DGRP is segregating for the African *D. melanogaster* CHC phenotype. There was also evidence that the DGRP contains genetic variants that suppress the African phenotype. These results provide a basis for additional genetic and biochemical studies to determine the molecular and biochemical functions of these candidate genes.

From the GWAS conducted on DGRP trait variation for this dissertation, similar trends have arisen. First, the genetic architecture of variation for each trait was more complex in the DGRP population than in populations used in previous gene mapping studies. Second, associated genetic variants were located in and near both known and novel candidate genes. These results provide a proof-of-principle that the DGRP population can be used to map genetic variants associated with natural phenotypic variation. Third, associated genetic variants are often in or near genes with pleiotropic effects. A majority of the genetic variants identified in my studies were located in intronic or intergenic regions. These results implicate *cis*-regulatory evolution, which likely limits negative pleiotropic effects, as a major contributor to phenotypic variation within the DGRP population. Future studies should characterize desiccation resistance in the DGRP and conducting GWA analyses. The results could represent genetic cross-links between CHC composition and pigmentation traits.
Quantitative Genetics of Cuticular Traits in *Drosophila melanogaster*

by

Lauren Marie Dembeck

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Genetics

Raleigh, North Carolina

2015

APPROVED BY:

_______________________________
Trudy F. C. Mackay
Committee Chair

_______________________________
Robert R. H. Anholt

_______________________________
Coby Schal

_______________________________
Nadia D. Singh

_______________________________
Eric A. Stone
DEDICATION

To my family, my biology teachers who were not afraid to teach me about evolution, and myself
I am fascinated by the immense diversity and peculiarity of life on Earth. In order to understand how complex traits evolve, potentially resulting in richer biodiversity, we must identify the genetic variants that underlie variation in those traits. That understanding led me to my Ph.D. studies using the *Drosophila melanogaster* Genetic Reference Panel under the guidance of Prof. Trudy Mackay at North Carolina State University and W. M. Keck Center for Behavioral Biology. I also studied mimicry, mate choice, and the speciation continuum in *Heliconius* butterflies with Dr. W. Owen McMillan at the Smithsonian Tropical Research Institute in Panama. In 2008, I graduated with a B.S. in Biology and B.A. in Foreign Languages from West Virginia University. While there, I worked with Dr. David A. Ray using accumulated genetic variation in retrotransposons to reconstruct the phylogenetic relationships of crocodiles, alligators, and gharials. I was fortunate to continue this research as a McNair Scholar through the Ronald E. McNair Post-Baccalaureate Achievement Program at WVU.

Natively, I come from Wallace, WV. During my youth, I trained to be a skilled outdoorswoman, fishing, catching crawdads and salamanders, climbing trees, shooting inanimate objects, driving vehicles through large mud holes, and rescuing kittens. These days, I enjoy exploring nature and meeting new organisms, yoga, knitting, photography, and cuddling with my feline companions Dennis and Merriam.

The studies of *Drosophila* presented in this dissertation have been accompanied by my continued growth as a researcher and as myself. I look forward to my continued exploration of this world and the things unknown.
ACKNOWLEDGMENTS

I am thankful to my partner and husband Casey James Galvin. He has enabled me to become a more thoughtful and thorough researcher and a more consistent and practiced writer. I am fortunate to spend my time with someone who is likeminded and just different enough to be wonderfully complementary to me. I look forward to continuing our scientific, artistic, and worldly endeavors.

I am grateful to Drs. Trudy F. C. Mackay (and Robert R. H. Anholt) for taking me in and offering me continued support and guidance during my Ph.D. I am very fortunate to have had the opportunity to work so closely with and learn from such a phenomenal scientist. I admire and respect Dr. Mackay and all of her scientific efforts and discoveries.

I am indebted to Dr. Anholt for his mentorship and for teaching me to dazzle 'em with style. If I ever give a good presentation, it is because of Robert's efforts. Robert also taught me many lessons about fine dining and selecting slightly better than average wines.

I owe a special thanks to each of my committee members. Dr. Nadia Singh has been an indispensable mentor to me. I cannot thank her enough for her contribution to my personal and professional development. Her encouragement and advice helped me assemble this dissertation and sometimes, reassemble myself. Dr. Coby Schal has been a mentor and collaborator throughout my entire graduate career. He is a model scientist and one I aspire to emulate. Dr. Eric Stone makes me feel like I could learn anything from him. His unwavering patience and willingness and ability to parse ideas to their most basic components are unlike any other professor. I also thank Dr. Jim Mahaffey for his help and guidance during the early stages of my program and his continued support (and tea).

To each of the following, I am thankful: North Carolina State University, the former Genetics Department (or the Department Formerly Known As Genetics) and now Biological Science Department Genetics Program, the NCSU Graduate School and Graduate Student Support Plan, Julieta Sherk for kindly serving as my graduate school representative, the College of Agriculture and Life Sciences, the College of Sciences, Katalin Böröczky, Aki Yamamoto, Ayako Katsumata, Richard Lyman, Wen Huang, Rachel Lyman, Ahkila Boyina, Hannah Moyles, Steven West, Rick Santangelo,
Faye Lawrence, Mike Magwire, Gunjan Arya, Lavanya Turlapati, Pat Estes, Sam To, Lenovia McCoy, Fly Food Crew (Alden, Thomas, and Nathan), Mary Anna Carbone, Carlos Botero, Tanya Morozova, Shilpa Swarup, Shanshan Zhou, Megan Garlapow, John Shorter, Julien Ayroles, Sarah Kocher, Beth Ruedi, Melissa Robbins, Jeff Hawley, Lanakila Alexander, Stephanie Curtis, Betty Gardner, Suzanne Quick, Franda Sagcal, David Threadgill, Jeff Thorne, Owen McMillan, Brian Counterman, Heather Hines, Moises Albanto, Stephanie Rusza, my instructors at NCSU and WVU, my fellow graduate students, David Ray, Keith Garbutt, WVU Eberly College of Arts and Sciences, WVU McNair Scholars Program, the WVU Alumni Association Loyalty Permanent Endowment Fund, the James V. Morris Scholarship Fund, Mr. J. V. Morris for his support and wonderful letters over the years, the J.P. Morgan-Chase/ Tom Stockert Foundation, the WV PROMISE scholarship program, the Greater Kanawha Valley Foundation, and Mr. Steve Gibson and many other wonderful teachers I had.

I also acknowledge and appreciate the help and love of my dear friends Terry Campbell and Laura Duncan. They were always there for me, scientifically and personally. My feline companions Dennis and Merriam have also offered me their unending love and comfort throughout this period of my life. I am grateful to Carolina Tiger Rescue for giving me the privilege to work as an animal care volunteer. It has fulfilled a dream of mine to study and work with big cats and was an honor to meet, know, and serve all of the magnificent animals at the Rescue. Thank you so very much.

I am forever thankful to my father Kevin for his never ending sacrifices, his encouraging me to continue my (at least, formal) education, and for drilling me to "speak clearly, slowly, and distinctly." I am also grateful to my mother Lisa for her love and support and for sharing with me her positive attitude and optimism toward life and kindness to others. I am thankful to my sister Melissa for being my best friend and always listening and helping me keep my head on straight. I thank my sister Andrea for her hospitality, encouragement, and lots of fun times. To my nieces and nephews, you all have warmed my heart and given me more reasons to cherish life and smile. I thank the rest of my family for their support, love, and understanding (especially mamma, thank you for always being so sweet). I have missed you all.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................................. ix

LIST OF FIGURES ................................................................................................................................. x

CHAPTER 1. INTRODUCTION ................................................................................................................ 1

THE DROSOPHILA MELANOGASTER GENETIC REFERENCE PANEL ......................................................... 1

BODY PIGMENTATION............................................................................................................................ 2

CUTICULAR HYDROCARBONS .............................................................................................................. 3

CHCs IN SOCIAL SETTINGS .................................................................................................................. 4

CHC BIOSYNTHESIS AND GENETICS................................................................................................... 4

REFERENCES ........................................................................................................................................... 6

CHAPTER 2. GENETIC ARCHITECTURE OF ABDOMINAL PIGMENTATION IN DROSOPHILA MELANOGASTER.... 11

ABSTRACT ............................................................................................................................................... 11

AUTHOR SUMMARY .............................................................................................................................. 12

INTRODUCTION ..................................................................................................................................... 12

RESULTS .................................................................................................................................................. 13

Quantitative genetics of pigmentation ................................................................................................. 13

Genome-wide association analyses ..................................................................................................... 16

Variance in pigmentation explained by top variants ............................................................................. 20

Comparison with previous studies ....................................................................................................... 20

Validation of candidate genes ............................................................................................................. 21

DISCUSSION .......................................................................................................................................... 27

MATERIALS AND METHODS ............................................................................................................... 30

Drosophila stocks and phenotyping ...................................................................................................... 30

Statistical and quantitative genetic analyses .......................................................................................... 30

Genome-wide association analysis ........................................................................................................ 31

Variance in pigmentation explained by top variants ............................................................................. 31

Validation of candidate genes ............................................................................................................. 32
Dissection and photography ................................................................. 33
ACKNOWLEDGMENTS ............................................................................. 33
REFERENCES ......................................................................................... 33

CHAPTER 3. Pleiotropic Immune Genes Contribute to Natural Variation in Cuticle Color in Drosophila melanogaster ................................................................. 38
ABSTRACT .................................................................................................. 38
INTRODUCTION .......................................................................................... 38
RESULTS ...................................................................................................... 39
Quantitative genetics of coloration .......................................................... 39
Genome-wide association analyses ......................................................... 40
Functional validation studies ................................................................... 41
DISCUSSION .............................................................................................. 46
MATERIALS AND METHODS .................................................................... 47
Drosophila stocks and phenotyping .......................................................... 47
Abdominal cuticle dissection and imaging .............................................. 47
Statistical and quantitative genetic analyses .......................................... 47
Genome-wide association analysis ......................................................... 48
Candidate gene validation experiments ................................................. 48
AUTHORS’ CONTRIBUTIONS ................................................................. 48
ACKNOWLEDGEMENTS ......................................................................... 49
REFERENCES ............................................................................................ 49

CHAPTER 4. The Genetic Architecture of Natural Variation in Cuticular Hydrocarbon Composition in Drosophila melanogaster ......................................................... 52
ABSTRACT .................................................................................................. 52
INTRODUCTION .......................................................................................... 53
RESULTS ...................................................................................................... 55
CHCs in the DGRP .................................................................................... 55
African female CHC phenotypes ............................................................. 56
CHAPTER 4. ANALYSIS OF CUTICULAR HYDROCARBONS

CHC correlations and principal components analyses ........................................................................ 60
GWA analyses .................................................................................................................................... 65
Functional validation of candidate genes ............................................................................................ 66
DISCUSSION ....................................................................................................................................... 75
Phenotypic characterization of CHC profiles ...................................................................................... 75
A segregating African CHC phenotype in a cosmopolitan population ............................................. 75
GWA analysis and functional assessment of candidate genes .............................................................. 76
MATERIALS AND METHODS ........................................................................................................... 77
Drosophila stocks and phenotyping .................................................................................................... 77
Quantification and identification of cuticular hydrocarbons .............................................................. 77
Statistical and quantitative genetic analyses ...................................................................................... 78
Genome-wide association (GWA) analyses ....................................................................................... 79
RNAi and mutant candidate gene validation experiments ................................................................. 79
Acknowledgements .......................................................................................................................... 80
Funding statement .............................................................................................................................. 80
Author contributions .......................................................................................................................... 80
REFERENCES .................................................................................................................................... 80

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS .................................................................. 85
CONCLUSIONS .................................................................................................................................. 85
FUTURE DIRECTIONS ......................................................................................................................... 86
REFERENCES .................................................................................................................................... 87
LIST OF TABLES

Table 2.1 Summary of candidate gene validation experiments. ..................................................23
Table 4.1 Cuticular lipids identified by GC-MS in DGRP males and females. ..........................57
Table 4.2 Phenotypes and In(3R)K genotypes for females from DGRP lines with functional Desat2 alleles ...............................................................................................................................................58
Table 4.3 Percent of CHC variation in the DGRP explained by PCs. ........................................63
LIST OF FIGURES

Figure 1.1 *Drosophila melanogaster* pigmentation pathway .................................................. 3
Figure 2.1 Natural variation in T5 and T6 pigmentation in the DGRP ........................................ 14
Figure 2.2 Natural variation in female abdominal pigmentation .............................................. 15
Figure 2.3 Genome-wide association analyses ........................................................................ 17
Figure 2.4 Pairwise linkage disequilibrium among the associated *bab1* variants ................... 19
Figure 2.5 Validation of pigmentation candidate genes using Exelixis insertion mutants and RNAi knockdown .............................................................................................................. 24
Figure 2.6 Female abdominal cuticles of Exelixis transposon insertion lines with significant effects on pigmentation .......................................................... 25
Figure 2.7 Female abdominal cuticles of with significant effects on pigmentation in lines with targeted RNAi knockdown constructs ............................................................ 26
Figure 3.1 Natural variation in T3 sclerotin and T6 melanin coloration in the DGRP .................. 42
Figure 3.2 Validation of coloration candidate genes using mutants and RNAi knockdown .......... 43
Figure 3.3 Representative images of mutant female abdominal cuticles with significant effects on T3 coloration .................................................................................................................. 44
Figure 3.4 Portions of T6 melanin and T3 sclerotin from lines expressing targeted RNAi knockdown constructs with effects on coloration ............................................................ 45
Figure 4.1 Representative male and female chromatograms from the DGRP ........................... 55
Figure 4.2 DGRP lines segregate for the female African CHC phenotype, *Desat2* allele, and *In3R(K)* inversion status ........................................................................................................ 59
Figure 4.3 MMC modules of DGRP female CHCs based on Spearman's rank correlation coefficients ($\rho$) ................................................................................................................................. 61
Figure 4.4 MMC modules of DGRP male CHCs based on Spearman's rank correlation coefficients ($\rho$) ................................................................................................................................. 62
Figure 4.5 Principal component biplots for PC1 and PC2 of DGRP CHCs ................................. 64
Figure 4.6 Summary of RNAi and mutant experiments for female CHCs ................................. 68
Figure 4.7 Summary of RNAi and mutant experiments for male CHCs ................................. 69
Figure 4.8 Example chromatograms of oenocyte-specific RNAi knockdowns and mutants – *CG13091* and *CG10097* ........................................................................................................... 70
Figure 4.9 Example chromatograms of oenocyte-specific RNAi knockdowns – *CG8680* and *CG30008* ................................................................................................................................. 72
CHAPTER 1. INTRODUCTION

Insects evolved from crustacean ancestors approximately 479 million years ago [1]. They arguably represent the most successful multi-cellular lineage on Earth making up approximately one million of the estimated 8.7 million species [2]. Early insects were confronted with two major challenges when moving from an aqueous to a terrestrial environment. First, in the open environment, they were exposed to higher levels of UV radiation from the sun and more fluctuating temperatures. Second, insect bodies have a high surface area to volume ratio due to their small size and are subject to rapid desiccation. Thus changes to the exoskeleton and in particular two cuticular traits, body pigmentation and cuticular hydrocarbon production, may have been critical for evolution within the class Insecta.

In order to understand how these traits evolve within and between species, we must first identify the genes underlying the production of these traits and genetic variants contributing to phenotypic variation [3]. For this dissertation I conducted genome-wide association studies to map the genetic architecture of natural variation in body pigmentation and cuticular hydrocarbon composition in the fruit fly Drosophila melanogaster.

THE DROSOPHILA MELANOGASTER GENETIC REFERENCE PANEL

Using D. melanogaster as a model organism offers many distinct advantages. The flies are easily maintained in controlled environmental conditions and have a rapid generation time. D. melanogaster has the largest repertoire of genetic tools available of any organism including deficiency and P-element mutagenesis stocks, RNA interference (RNAi) lines, balancer chromosomes, and classic visual genetic markers [4].

Within the past decade, advances in DNA sequencing technologies have made mapping the genetic architecture of variation in complex traits more tractable than ever. This enabled the development of the Drosophila melanogaster Genetics Reference Panel (DGRP) [5]. The DGRP is a living library of 205 wild-derived inbred lines of D. melanogaster for which full-genome sequences are available. In 2002, individual gravid females were collected at the Raleigh Farmers’ Market and the lines were generated by 20 generations of full-sib inbreeding of their progeny. The genetic variation within each line is minimized while variation present in the wild-population is captured between the lines. After sequencing, approximately 4.5 million single nucleotide polymorphisms (SNPs), insertion-deletion (indel) polymorphisms, and microsatellites were identified [5,6].

This panel and the segregating genetic variants allow the genetic basis of natural variation in complex traits to be mapped using a genome-wide association study (GWAS). GWAS are based on linkage disequilibrium (LD)—non-random segregation—between a polymorphic genetic marker and
genetic variants causing phenotypic variation in a trait. In *D. melanogaster*, LD breaks down rapidly over short physical distances due to recombination during meiosis. In the DGRP, linkage disequilibrium breaks down within ~150-200 base pairs; therefore, any genetic variants associated with phenotypic variation are likely to be causal quantitative trait loci (QTL), nucleotides (QTN), or very close to the causal locus [5,6]. For a given trait, this provides enough power to detect intermediate frequency variants with relatively small to large effects. Thus the DGRP serves as a community resource in which an inestimable number of phenotypes can be measured and provides a powerful resource to conduct GWAS.

**BODY PIGMENTATION**

Pigmentation is a conspicuous trait that is often variable both within and between species [7–9]. Within a species, it is important for thermoregulation, UV protection, species recognition, mate choice, mimicry, and camouflage [7]. The fitness of an organism is often dependent upon proper coloration and thus it is subject to adaptive evolution and sexual selection. Many of the genes that contribute to insect coloration are known to be pleiotropic for many other traits including development, immune responses, and behavior [10–12].

The melanin and sclerotin biosynthesis pathway has been well-studied in *D. melanogaster* (Figure 1.1) [8,13]. Tyrosine is the starting substrate. Tyrosine hydroxylase encoded by *pale*, converts it into dihydroxyphenylalanine (DOPA) by adding a hydroxyl group. DOPA decarboxylase (*Ddc*) then decarboxylates DOPA into dopamine. DOPA and dopamine are then utilized for making four possible products: DOPA melanin, dopamine melanin, yellowish-tan N-β-alanyl dopamine (NBAD) sclerotin, and colorless N-acetyldopamine (NADA) sclerotin [7,13]. Though this biosynthetic pathway is well studied, there are many (40+) genes known to affect pigmentation in *D. melanogaster* and many of them have not been placed in this pathway or any parallel pathway [13].

In order to fill in gaps and link genes to enzymatic processes, I mapped the genetic architecture of natural variation in body pigmentation in the DGRP. I conducted two independent studies, chapters 2 and 3, on body pigmentation. The first focused on the proportions of the dorsal body segments known as tergites that are colored with dark melanin. The second assessed the intensity of the color of the dark melanin and yellowish-tan sclerotin on each of the tergites. These studies also allowed me to identify novel candidate genes not previously known to affect pigmentation.
Figure 1.1 *Drosophila melanogaster* pigmentation pathway. Gene names are in italicized text. Double arrows indicate segments with other unknown genes and reactions including phenol oxidases (PO). Melanin and sclerotin products are colored according to mature adult cuticle coloration.

**CUTICULAR HYDROCARBONS**

As previously mentioned, the evolution of cuticular hydrocarbon (CHC) production was necessary for insects to colonize terrestrial environments. CHCs are a blend of fatty acid derived waxes found on the cuticle [14]. They are produced by oenocytes, specialized subepidermal cells, and then transported through the hemolymph to the cuticle [15,16]. Their primary role is desiccation resistance [17]. There is no evidence for de novo CHC production in crustaceans and little evidence in other arthropods [18,19]. Thus CHC biosynthesis is a novel trait specific to insects among the Arthropoda.

The general model for insect desiccation resistance is that at moderate temperatures CHCs form a solid layer over the cuticle making a hydrophobic barrier preventing cuticular transpiration. With increasing temperatures, even within physiological ranges, some or all of the lipids melt, which creates a more permeable layer allowing rapid water-loss [20,21]. Methylbranched and desaturated CHCs render the lipid layer more porous because they cannot pack tightly. In contrast, saturated CHCs can align forming crystalline layers that are more protective. However, overall, the composition and distribution of the CHCs on the cuticle, their physical properties, and interactions with each other and the cuticle determine how well-insulated the insect is [17,18,20,22].
In many species, CHCs facilitate social interactions by conveying information such as species and nestmate identity, caste, age, and reproductive status [21,23]. They also serve as contact pheromones during mate choice [24,25]. Though *Drosophila* use a combination of all sensory systems during courtship and mating, CHC composition is particularly important in *D. melanogaster* mate choice [23]. As a result, CHCs have received attention for their potential role in establishing reproductive isolation and speciation [26–34].

In *D. melanogaster*, CHCs are sexually dimorphic. Males produce mostly short-chained alkanes and monoenes. 7-tricosene (7-C23:1) and 7-pentacosene (7-C25:1) are the primary and secondary male pheromones, respectively [28,35]. Males also produce an anti-aphrodisiac and aggregation pheromone called *cis*-vaccenyl acetate (cVA) [24,36,37]. It is transferred from the ejaculatory bulb to females during mating. cVA is not a CHC but it does leak from the genitalia to the posterior cuticle and can be detected in external body extracts of male flies [38].

Females produce a much more complex blend that also contains dienes, long-chained CHCs, and more methylbranched alkanes [28]. However, similar to males, they have two major pheromonal CHCs. The primary pheromone is 7,11-heptacosadiene (7,11-C27:2), and the secondary is 7,11-nonacosadiene (7,11-C29:2). It has been proposed that these tactile pheromones do not serve only as attractants but also as inhibitory pheromones. *D. melanogaster* females with little to no CHCs are more attractive to males, even conspecific males, than wild-type females [39].

Studies of female mate choice are becoming more prevalent [40–43]. The possible case of incipient speciation between *D. melanogaster* Cosmopolitan and African races provides a well-studied example [29,34,44–46]. African females produce the isomers, 5, 9-C27:2 and 5, 9-C29:2, as their primary pheromones and have reduced levels of 7,11-C27:2 and 7,11-C29:2. Though one may expect males to discriminate against females of the other race, males show no preference. Instead, African females discriminate against Cosmopolitan males; however, the cause of this discrimination is unclear since Cosmopolitan and African males do not differ in their CHC profiles.

**CHC BIOSYNTHESIS AND GENETICS**

*D. melanogaster* has served as a powerful model system in which to study the biosynthesis of CHCs and the genetic variation contributing to CHC differences within and between species. One study has comprehensively analyzed CHCs using a quantitative genetics approach to identify genomic regions contributing to natural variation in *D. melanogaster* CHCs [47]. A panel of 144 isogenized lines derived from two gravid, wild females, the Winter's Lines, was studied. Variation in all detectable CHCs was analyzed among these lines and correlated with polymorphic retrotransposons. This approach allowed large, often overlapping, chromosomal regions contributing to CHC variation to be mapped; however, it did not have the resolution to identify individual genes.
The *Desat* locus in *D. melanogaster* was the first description of desaturase genes in insects [48]. The locus consists of two genes, *Desat1* and *Desat2*. *Desat1* is expressed in both sexes and has been the most well studied CHC biosynthesis gene. The enzyme is a delta-9 desaturase that leads to the synthesis of palmitoleic acid, a precursor of omega-7 fatty acids and 7-unsaturated hydrocarbons [49,50]. It has also been shown to have pleiotropic effects on male perception of sex pheromones which may be attributed to alternative splice variants [27,50]. In addition to *Desat1*, Cyp4G1, a NADPH-cytochrome P450 reductase, acts in both sexes; it was shown to be the oxidative decarbonylase that converts long-chained aldehydes to hydrocarbons [51].

*Desat2* has received considerable attention as a "speciation gene" and for its role in the divergence of African and Cosmopolitan races of *D. melanogaster* mentioned above [52]. It is reportedly inactive in Cosmopolitan females with the 7, 11-diene pheromones and active in African and some Caribbean populations leading to the production of the 5, 9-dienes [29,53,46,54]. Females with an active *Desat2* also exhibit a strong bias against Cosmopolitan males suggesting a pleiotropic role similar to that of *Desat1* [44,34,45,26].

The question still remains as to how the precursors for the 5, 9- and 7, 11-dienes change in the two races and what other genes may be involved. *Desat1* desaturates palmitic acid (16:0) to palmitoleic (16:1 Δ9). *Desat2* desaturates myristic acid (14:0) to myristoleic acid (14:1 Δ9). These intermediate products are then acted upon by other enzymes to produce the final pheromonal CHCs in both races. However, in Cosmopolitan *D. melanogaster*, there is more palmitic acid (49% of fatty acids synthesized) available than myristic acid (29%) suggesting that either *Desat2* is able to desaturate the myristic acid before it can be elongated to palmitic acid or the fatty acid synthase produces more myristic acid than palmitic acid in African females [55]. The latter has not been investigated.

Other studies of *D. melanogaster* female specific or biased CHC production have uncovered some of the key genes responsible, *DesatF* and *elongase F* (*eloF*) [48]. *DesatF* is responsible for creating the second double bond in the dienes of females [56,57]. *eloF* has been implicated in the production of long-chain CHCs; *eloF* RNAi knockdown females displayed an increase in *n*-C23 and *n*-C25 and a corresponding decrease in the pheromonal C27 and C29 dienes [58].

Studies focusing on few CHCs using distant and distinct populations have mapped individual genes contributing to male *D. melanogaster* CHC differences. Genomic regions called *small monoene quantities* (*smoq*) and *seven pentacosene* (*sept*) were shown to contribute to variation in the proportions of 7-C23:1 and 7-C25:1 [59]. *nerd*, identified through a mutagenesis screen, drastically reduced 7-C23:1 production and altered courtship behavior [60].

Outside of the aforementioned genes, no other CHC biosynthesis genes have been identified. In chapter 4, I present a study of the genetic architecture of natural variation in cuticular
hydrocarbon composition in DGRP males and females. I conducted a genome-wide association study and tested 24 candidate genes for effect on CHC composition. I also describe evidence for an epistatic interaction between a Desat2 indel and other variants in the DGRP. The identified genes provide a foundation upon which to generate new hypotheses that may offer insight into insect evolution and diversification.

REFERENCES


CHAPTER 2. GENETIC ARCHITECTURE OF ABDOMINAL PIGMENTATION IN DROSOPHILA MELANOGASTER

Lauren M. Dembeck¹,²,³, Wen Huang¹,²,³, Michael M. Magwire⁴, Faye Lawrence¹, Richard F. Lyman¹,²,³, and Trudy F. C. Mackay¹,²,³

¹Department of Biological Sciences, ²Program in Genetics and ³W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695-7614, USA
⁴Syngenta Biotechnology, 3054 E. Cornwallis Road, Research Triangle Park, NC 27709 USA
DOI: 10.1371/journal.pgen.1005163

ABSTRACT

Pigmentation varies within and between species and is often adaptive. The amount of pigmentation on the abdomen of Drosophila melanogaster is a relatively simple morphological trait which serves as a model for mapping the genetic basis of variation in complex phenotypes. Here, we assessed natural variation in female abdominal pigmentation in 175 sequenced inbred lines of the Drosophila melanogaster Genetic Reference Panel, derived from the Raleigh, NC population. We quantified the proportion of melanization on the two most posterior abdominal segments, tergites 5 and 6 (T5, T6). We found significant genetic variation in the proportion of melanization and high broad-sense heritabilities for each tergite. Genome-wide association studies identified over 150 DNA variants associated with the proportion of melanization on T5 (84), T6 (34), and the difference between T5 and T6 (35). Several of the top variants associated with variation in pigmentation are in tan, ebony, and bric-a-brac1, genes known to affect D. melanogaster abdominal pigmentation. Mutational analyses and targeted RNAi-knockdown showed that 17 out of 28 (61%) novel candidate genes implicated by the genome-wide association study affected abdominal pigmentation. Several of these genes are involved in developmental and regulatory pathways, chitin production, cuticle structure, and vesicle formation and transport. These findings suggest that both pleiotropy and genetic heterogeneity contribute to the genetic architecture underlying variation in body pigmentation. Variation in these novel candidates may serve as targets for adaptive evolution and sexual selection in D. melanogaster.
Body pigmentation contributes to the spectacular biodiversity present in nature and mediates mate choice, mimicry, and physiological functions such as thermoregulation and UV resistance. Thus, pigmentation is a significant contributor to fitness. In order to understand how complex traits such as pigmentation evolve, we must first identify the genetic variants underlying phenotypic variation. We used the *Drosophila melanogaster* Genetic Reference Panel, a wild derived population of fully sequenced inbred fly lines, to identify the contributions of both known and novel genetic variants to natural variation in abdominal pigmentation in female flies. Our results show that genetic variation within many biological pathways contributes to variation in *D. melanogaster* pigmentation.

**INTRODUCTION**

Body pigmentation is a conspicuous trait that is variable within species, giving rise to natural variation, polyphenism and sexual dimorphism [1–4]. It also varies between species, contributing to species recognition, mate choice, thermoregulation, protection (warning signals), mimicry, and crypsis [5–7]. Changes in pigmentation are often adaptive and vital to the fitness of the organism [5,6]. Not only is body pigmentation ecologically relevant, in *Drosophila* it is a relatively simple and easily measured phenotype to study the genetic architecture of natural variation in complex traits [2,7–10]. Each tergite of female *D. melanogaster* generally has a stripe of dark coloration (melanin) on a lighter tan background (sclerotin). During pre- and post-ecdysis, the epidermal cells underlying the cuticle secrete tyrosine-derived catecholamines into the cuticle for sclerotinization and melanization [11,12]. The melanin/sclerotin biosynthetic pathway and its underlying genetic basis have been well studied. However, many of the genes known to affect *D. melanogaster* pigmentation do not form part of this pathway or any parallel pathway [5,13]. Furthermore, the genes that lead to natural variation in body pigmentation are not necessarily the same genes that are directly involved in the biosynthesis of melanin and sclerotin. By mapping the genetic basis of natural variation in body pigmentation, we may discover new genes affecting pigment biosynthesis as well as regulatory regions that determine when and where pigmentation will develop [3,13].

We used the *D. melanogaster* Genetic Reference Panel (DGRP) to perform a genome-wide association (GWA) study of natural variation in the proportion of melanization on female abdominal tergites 5 and 6. The DGRP consists of 205 sequenced inbred lines derived from a single North American population, facilitating GWA analyses for quantitative traits when all genetic variants are known. Local linkage disequilibrium (LD) in the DGRP is low and thus favorable for identifying candidate genes and even causal polymorphisms [14,15]. We identified single nucleotide polymorphisms (SNPs) affecting three genes previously known to contribute to variation in abdominal pigmentation, *bric-à-brac 1 (bab1)*, *tan (t)*, and *ebony (e)*. However, we also identified novel candidate genes and showed that these contribute to abdominal pigmentation using mutations and RNAi knock-
down constructs. Many of these novel genes affect other well-studied pathways and phenotypes, such as wing and bristle development, providing evidence for widespread pleiotropy and genetic heterogeneity. Four of the novel genes affecting pigmentation are computationally predicted genes with previously unknown functions. Based on their mutant or RNAi knockdown phenotypes, we have named them *pinstripe* (*pns, CG7852*), *triforce* (*tfc, CG9134*), *plush* (*ph, CG1887*), and *farmer* (*frm, CG10625*).

**RESULTS**

*Quantitative genetics of pigmentation*

We characterized natural variation in the proportion of melanization of tergites 5 (T5) and 6 (T6) in females for 175 DGRP lines (Fig 2.1, Fig 2.2). Averaged across all lines, the mean pigmentation scores are 1.44 for T5 and 2.55 for T6 (Fig 2.2A). There is significant genetic variation in pigmentation among lines for both tergites ($P_{T5} = 4.68 \times 10^{-48}$ and $P_{T6} = 6.65 \times 10^{-96}$), with broad sense heritabilities ($H^2$) of $H^2_{T5} = 0.66$ and $H^2_{T6} = 0.88$. The phenotypic ($r_{p(T5,T6)} = 0.63 \pm 0.059$) and genetic ($r_{g(T5,T6)} = 0.72 \pm 0.053$) correlations (± standard error) between the tergites for proportion of pigmentation are high but significantly different from unity, suggesting they have different genetic bases (Fig 2.2B). The high broad sense heritabilities for abdominal pigmentation traits provide a favorable scenario for GWA studies.
Figure 2.1 Natural variation in T5 and T6 pigmentation in the DGRP. Images display one half of the fly abdominal cuticle, split along the dorsal midline. Pigmentation scores are given for T5 (upper) and T6 (lower). DGRP lines are denoted by DGRP_XXX in the upper left corner of each image. (A) Variation in T5. (B) Variation in T6. (C) Variation in spatial patterning of pigmentation.
Figure 2.2 Natural variation in female abdominal pigmentation. (A) T5 (gray) and T6 (red). DGRP lines are in order from least to most pigmentation on T6. (B) Scatter plot of T5 and T6 line means.
**Genome-wide association analyses**

We performed genome-wide association analyses on the proportion of T5 and T6 melanization to identify genomic regions harboring variants contributing to natural variation in female abdominal pigmentation. The DGRP lines vary in *Wolbachia* infection status and karyotype for several common polymorphic inversions. We did not find significant associations of *Wolbachia* infection (\(P_{T5} = 0.58\) and \(P_{T6} = 0.92\)) nor inversion karyotype on T5 or T6 pigmentation; however, the difference in pigmentation between T5 and T6 was significantly affected by \(\ln(2L)t\) (\(P = 0.04\)) and \(\ln(2R)NS\) (\(P = 0.01\)). For each GWA analysis, we used both a mixed model that accounted for any effects of *Wolbachia*, inversions, and cryptic relatedness and a regression model that corrected for all of the aforementioned effects except for cryptic relatedness [15]. Combining all of these models, we identified a total of 155 variants associated with pigmentation for any trait at a nominal reporting threshold of \(P < 10^{-5}\). Of these, 84 were associated with T5 pigmentation, 34 with T6 pigmentation, 28 with the average of T5 and T6, and 35 with the difference in pigmentation between T5 and T6. A total of 84 candidate genes were implicated by these associated variants. Since variants associated with the average of the two posterior tergites were largely the same as those associated with either T5 or T6 alone, we focus our subsequent analyses on T5, T6 and the difference between them (Fig 2.3).

Among the genes harboring SNPs associated with variation in abdominal pigmentation, we find genes with well documented effects on pigmentation (*t*, *e*, *bab1*); *osa*, a transcription factor recently shown to affect pigmentation; and a large group of novel candidate genes [2,9,16]. The identification of *t*, *e*, and *bab1* as prominent contributors to variation in abdominal pigmentation instills confidence in the efficacy of our GWA analyses, as described below.

Only a few variants exceeded a strict Bonferroni correction for multiple tests (\(P = 2.64 \times 10^{-8}\)): a SNP 41 bp upstream of *Gr8a* and 528 bp downstream of *CG15370* – the cis-regulatory region of *t* – in the T6 and average of T5 and T6 analyses (X_9121129_SNP); and two SNPs in the first intron of *bab1* in the analysis of the difference between T5 and T6 (3L_1084990_SNP and 3L_1084199_SNP). The three SNPs that achieved Bonferroni significance levels were all at intermediate frequency and had moderately large effects. The minor allele of the polymorphism in the *t* cis-regulatory element (CRE) was associated with reduced pigmentation, while the minor alleles of the *bab1* intronic polymorphisms were both associated with increased pigmentation in T6.
Figure 2.3 Genome-wide association analyses. Results are depicted for T5, T6, and the T5-T6 difference. A nominal $P \leq 10^{-5}$ is indicated with a red line for each trait. The triangular heat map depicts the degree of LD, $\hat{r}^2$, between variants. The five major chromosome arms are delineated by the black lines. Red corresponds to complete LD and blue to absence of LD. The upper panels show the mixed model significance threshold (-log10($P$)) and the effect size in phenotypic standard deviation units ($a/\sigma_p$) for each trait. The minor allele frequency (MAF) is shown on the bottom panel.
Although the other variants do not reach individual Bonferroni-corrected significance levels, quantile-quantile plots indicate a systematic departure from random expectation below $P < 10^{-5}$, justifying our choice of this reporting threshold and suggesting that the top associations are enriched for true positives. Indeed, the SNP in the $t$ CRE that reached Bonferroni significance in the T6 analysis was also significant in the T5 analysis at the more lenient reporting threshold, and two additional polymorphisms in the $t$ CRE were significant at $P < 10^{-5}$: $X_{9121177}$ SNP in the T5 and T6 analyses, and $X_{9121094}$ SNP in the T6 analysis.

These data also highlight the importance of $bab1$ with respect to female abdominal pigmentation: we found a total of 21 polymorphisms (20 SNPs, one indel) in the first intron of this gene that are associated with natural variation in pigmentation in one or more analyses (Fig 2.4). One $bab1$ SNP is unique to the T5 analysis, six $bab1$ SNPs are common between the T6 and T5 – T6 difference analysis, and the remaining $bab1$ variants are unique to the difference in pigmentation between T5 and T6. Twelve of the $bab1$ variants are located within the minimal functional cis-regulatory regions as reported by REDfly or within other transcription factor binding sites [17]. Three SNPs ($3L_{1084990}$ SNP, $3L_{1085137}$ SNP, and $3L_{1085230}$ SNP) are located in the $bab1$ middle dimorphic element which contain binding sites for the transcription factors $caudal$ ($cad$) and $dl$ ($dorsal$) [3]. All of the polymorphisms segregating in $bab1$ associated with pigmentation have minor allele frequencies ranging from 0.22 to 0.49 and moderate effects. Interestingly, the direction of the effects is both positive (the minor allele is associated with reduced pigmentation) and negative (the minor allele is associated with increased pigmentation), such that variants in the $dl$ and $cad$ cis-regulatory modules have positive effects while those in the latter regions of the intron have largely negative effects (the exceptions are $3L_{1093297}$ SNP in the latter region of the intron and $3L_{1099962}$ SNP in the T5 GWAS, which have positive effects). The functionality of these $bab1$ CREs has been thoroughly investigated [3]. However, similar to the results of Bickel et al. [18], nine of the $bab1$ variants from this study are in regions outside of the known cis-regulatory regions. These variants may indicate the presence of a not-yet-described regulatory element, or the structure of the regulatory elements in this region may be more complex than previously thought (Fig 2.4).
Figure 2.4 Pairwise linkage disequilibrium among the associated *bab1* variants. Heat map of LD ($r^2$) of the 21 *bab1* variants. Individual variants, cis-regulatory regions and transcription factor binding sites are labeled on the diagonal.
A majority of the variants associated with variation in pigmentation are located within intronic or intergenic regions, suggesting they could affect gene regulation. In support of this hypothesis, we found many of these variants are located in annotated regulatory sites. In total, variants associated with pigmentation were located in 24 different transcription factor binding sites (TFBS, each of which contain numerous variants), 17 cis-regulatory modules (CRM), 1 polycomb response element (PRE), and 31 hot spot analysis sites (HSA; where one or more 41 tested TFs bind to a given site). TFBS for *dl* and *cad* are the most frequent of all TFBS, containing 28 and 22 associated variants, respectively. Two intergenic TFBS for *bab1* (FBsf0000214860 and FBsf0000214320) were tagged by 3R_25139342_SNP, 3R_25139132_SNP, and 2R_16793853_SNP. A few variants are located in more than one regulatory site.

**Variance in pigmentation explained by top variants**

We asked what fraction of the total broad sense heritability was explained by variants in *bab1*, *t* and *e* using stepwise regression to select the top associations for pigmentation genes. The $R^2$ from these models for each trait gives the heritability explained by the known genes. These loci explain 25.62%, 37.55%, 31.17% and 36.58% of the heritability for T5, T6, and the average and difference of T5 and T6, respectively; consistent with the intermediate allele frequencies and large effects of their top associated variants. Next, we used genomic best linear unbiased prediction (GBLUP) to estimate the total variance explained by all top variants. All variants explain 59.77%, 34.32%, 47.44% and 51.61% of the heritability for T5, T6, and the average and difference of T5 and T6, respectively. With the exception of T6, for which most of the variance is explained by the known pigmentation genes, substantial additional variance in proportion of pigmented cuticle is contributed by variants in novel genes. Finally, we estimated the fraction of heritability explained for each variant as well as the fraction of heritability explained after accounting for the variance explained by the pigmentation genes. On average, the variants in novel candidate genes explained an additional 7.3% (T5), 4.5% (T6), 5.8% (average of T5 and T6) and 2.8% (difference between T5 and T6) of the heritability.

**Comparison with previous studies**

*e*, *t*, and *bab1* have been associated with variation in *D. melanogaster* female pigmentation in other populations [2,9,16]. We compared the top variants in these genes in our analyses with those from prior studies [9,16,19]. Three of the four SNPs identified by Bastide *et al.* [9] are identical to the three *t* CRE SNPs associated with our T5 and T6 analyses (*X*_9121094_SNP, *X*_9121129_SNP, and *X*_9121177_SNP). The *bab1* SNP identified in the Bastide *et al.* [9] study did not overlap with our results nor those of Bickel *et al.* [18]. None of the top *bab1* variants in this study were significant in the study of Bickel *et al.* [18], although three of our significant variants were also polymorphic in the Bickel data set (3L_1085788_SNP, 3L_1086799_SNP, and 3L_1086802_INS). Both the *e* CRE SNPs
associated with pigmentation in the DGRP T6 analysis (3R_17063120_SNP) and the Bastide et al. study [9] (3R_17064232_SNP) are located within the CRE regulating e expression in the haltere (e_C [19]). The SNP at 3R_17064232 was also reported in the Pool and Aquadro study of light and dark African D. melanogaster [16]. The concordance among these datasets indicates that the haltere regulatory element may also control expression in the abdomen and warrants further investigation.

Validation of candidate genes
We selected 30 novel candidate genes based on the GWA results for functional validation using mutant alleles and RNAi knockdown. We phenotyped Exelixis insertion lines [20] and RNAi knockdown lines [21] with their appropriate controls for the proportion of melanization on T5 and T6 as done for the DGRP. Wherever possible we tested both mutant and RNAi lines for the same gene as independent forms of validation. We used three GAL4 drivers for the UAS-RNAi lines. tubulin-GAL4/TM3, Sb (tub-GAL4) and ubiquitin-GAL4/Cyo (ubi-GAL4) are ubiquitously expressed, while the pannier driver, y¹ w¹¹¹, P[w¹+mW.hs]=GawB)pnr^MD237/TM3, P[w¹+mC]=UAS-y.C)MC2, Ser¹ (pnr-GAL4), has restricted expression in the midline [22]. The use of the pnr-GAL4 driver adds a spatial component to the validation experiments and allows for more precise testing of the candidate genes.

As positive controls, we also tested RNAi constructs for e and t.

We evaluated 15 Exelixis transposon insertions in candidate genes for effects on pigmentation (See Methods; Fig 2.5A). Six of these mutations affected the proportion of melanization on T5 (P < 0.0001 for all significant mutations): CG9134^e00088, CG7852^f04511, Exchange factor for arf6 (Efa6^f03476), Fish-lips (Fili^f04573), and Glucose transporter 1 (Glut1^d05758) showed increased melanization; and krotzkopf verkehrt (kkv^f06225) showed decreased melanization (Fig 2.5A, Fig 2.6).

Twelve of the mutations affected the proportion of melanization on T6 (P < 0.0001 for all significant mutations). CG9134^e0008, Efa6^f03476, Fili^f04573, and Glut1^d05758 showed increased melanization; and CG10625^e01211, CG7852^f04511, division abnormally delayed (dally^f01097), kayak (kay^f02003), kkv^f06225, klersicht (klar^f05910), locomotion defects (loco^f09879), and Sucb^f01940 showed decreased melanization (Fig 2.5A, Fig 2.6). CG7852^f04511 had increased pigmentation on T5 and decreased pigmentation on T6 (Fig 2.6C). CG33298^f01067a, multiple wing hairs (mwh^f01620), and Kinesin-like protein at 61F (Klp61F^f02870) were not significantly different from the control. Efa6^f03476 also has a light and somewhat elongated thoracic trident; this thoracic pigmentation is completely absent in the control flies.

Of the 28 candidate genes, 26 were available as RNAi knockdown constructs. We crossed all of these constructs to the pnr-GAL4 driver, and obtained viable female progeny from all crosses except for kkv and Fili. We found that seven of these knockdown mutations affected the proportion of melanization of T5 and/or T6 (Fig 2.5B, Fig 4.7A-I, P < 0.0001 in all cases). Efa6, klar, and Klp61F knockdowns increased the proportion of dark melanin on T6; buttonless (btn) and CG7852 decreased
it. Knockdown of roughoid (ru) and sinuous (sinu) showed decreases in pigmentation for both T5 and T6.

We also crossed the 26 UAS-RNAi constructs to an ubiquitously expressed tub-GAL4 driver, and found that 11 (42%) were lethal in both sexes, suggesting pleiotropic effects on vital functions: Vesicular monoamine transporter (Vmat), Klp61F, CG7852, CG1887, klar, ru, sinu, Nedd2-like caspase (Nc), kkv, kay, and CG42340 (Table 2.1). Of the 15 UAS-RNAi/tub-GAL4 knockdown alleles available for testing, six had significant \( P < 0.0001 \) effects on pigmentation. Knockdowns of btn, CG10625, and CG9134 had decreased proportions of dark melanin on T5 and T6; Efa6 and loco knockdowns showed increases in pigmentation on both tergites (Fig 2.5C, Fig 4.7).

Next, we crossed the 11 UAS-RNAi constructs that were lethal when crossed to the tub-GAL4 driver to another ubiquitously expressed GAL4 driver, ubi-GAL4. Only two RNAi constructs were viable when driven by ubi-GAL4, CG1887 and klar, and both had significant \( P < 0.0001 \) effects on abdominal pigmentation (Fig 2.5D, Fig 4.7, and Table 2.1). The CG1887 knockdown showed a decrease in T5 melanization. Although T6 did not show a significant difference in the proportion of melanin in the CG1887 knockdown \( (P = 0.71) \), the dark melanin that is present is a light brown melanin that is only slightly darker than the adjacent sclerotin. The CG1887 knockdown flies have obvious qualitative differences in overall body coloration from the control. The cuticle as a whole is semi-transparent and its strength is compromised as it ruptures easily when probed with forceps. The third thoracic leg of these progeny is also malformed and bristle number and patterning is highly disrupted. All progeny of the cross die within 24 hours of eclosion; thus, pigmentation scoring was performed 8 hours after eclosion. The klar knockdown shows an increase of melanization on both tergites. Similar to the Efa6 mutant, this cross leads to a relative darkening of the thoracic trident compared to the surrounding cuticle. In summary, we found that 17 of the 28 candidate genes tested affected female abdominal pigmentation and that for 12 of these genes, both tergites are affected (Table 2.1).
Table 2.1 Summary of candidate gene validation experiments. NS = not significant,"- " = line not available or RNAi cross not tested. ↑ = increased pigmentation. ↓= decreased pigmentation.

<table>
<thead>
<tr>
<th>Candidate Gene</th>
<th>GWAS Association</th>
<th>Exelixis (T5/T6)</th>
<th>ExelixisRNAi Driver</th>
<th>ExelixisRNAi Driver</th>
<th>ExelixisRNAi Driver</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG33298</td>
<td>T6</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>NS/NS</td>
</tr>
<tr>
<td>Fish-lips (Fili)</td>
<td>T6</td>
<td>↑/↑</td>
<td>lethal (♀ only)</td>
<td>lethal (♀ only)</td>
<td>-</td>
</tr>
<tr>
<td>Vesicular monoamine transporter (Vmat)</td>
<td>T6</td>
<td>-</td>
<td>lethal</td>
<td>NS/NS</td>
<td>NS/NS</td>
</tr>
<tr>
<td>multiple wing hairs (mwh)</td>
<td>T6</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>NS/NS</td>
</tr>
<tr>
<td>Kinesin-like protein at 61F (Klp61F)</td>
<td>T6</td>
<td>NS/NS</td>
<td>lethal</td>
<td>NS/↑</td>
<td>lethal</td>
</tr>
<tr>
<td>CG9134</td>
<td>T6</td>
<td>↑/↑</td>
<td>↓/↓</td>
<td>NS/↓</td>
<td>-</td>
</tr>
<tr>
<td>CG7852</td>
<td>T6, T5-T6</td>
<td>↑/↓</td>
<td>lethal</td>
<td>NS/↓</td>
<td>-</td>
</tr>
<tr>
<td>CG1887</td>
<td>T6</td>
<td>-</td>
<td>lethal</td>
<td>NS/NS</td>
<td>↓/NS</td>
</tr>
<tr>
<td>klarsicht (klar)</td>
<td>T6</td>
<td>NS/↓</td>
<td>lethal</td>
<td>NS/↑</td>
<td>↑/↑</td>
</tr>
<tr>
<td>Glucose transporter 1 (Glut1)</td>
<td>T6</td>
<td>↑/↑</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Exchange factor for Arf 6 (Efa6)</td>
<td>T6</td>
<td>↑/↑</td>
<td>↑/↑</td>
<td>NS/↑</td>
<td>-</td>
</tr>
<tr>
<td>buttonless (btn)</td>
<td>T6</td>
<td>-</td>
<td>↓/↓</td>
<td>NS/↓</td>
<td>-</td>
</tr>
<tr>
<td>rougoid (ru)</td>
<td>T5</td>
<td>-</td>
<td>lethal</td>
<td>↓/↓</td>
<td>lethal</td>
</tr>
<tr>
<td>CG10625</td>
<td>T5</td>
<td>NS/↓</td>
<td>↓/↓</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>sinuous (sinu)</td>
<td>T5</td>
<td>-</td>
<td>lethal</td>
<td>↓/↓</td>
<td>lethal</td>
</tr>
<tr>
<td>Sucb (Sucb)</td>
<td>T5</td>
<td>NS/↓</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>division abnormally delayed (dally)</td>
<td>T5</td>
<td>NS/↓</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>CG32052</td>
<td>T5</td>
<td>-</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>Nedd2-like caspase (Nc)</td>
<td>T5</td>
<td>-</td>
<td>lethal</td>
<td>NS/NS</td>
<td>lethal</td>
</tr>
<tr>
<td>Ceramide kinase (Cerk)</td>
<td>T5</td>
<td>-</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>krotzkopf verkehrt (kkv)</td>
<td>T5</td>
<td>↓/↓</td>
<td>lethal</td>
<td>lethal</td>
<td>lethal</td>
</tr>
<tr>
<td>CG15803</td>
<td>T5</td>
<td>-</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>locomotion defects (loco)</td>
<td>T5</td>
<td>NS/↓</td>
<td>↑/↑</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>TweedleC (TwdlC)</td>
<td>T5</td>
<td>-</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>kayak (kay)</td>
<td>T5</td>
<td>NS/↓</td>
<td>lethal</td>
<td>NS/NS</td>
<td>lethal</td>
</tr>
<tr>
<td>CG1340</td>
<td>T5</td>
<td>-</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>CG42594</td>
<td>T5</td>
<td>-</td>
<td>NS/NS</td>
<td>NS/NS</td>
<td>-</td>
</tr>
<tr>
<td>CG42340</td>
<td>T5</td>
<td>-</td>
<td>lethal</td>
<td>NS/NS</td>
<td>lethal</td>
</tr>
</tbody>
</table>
Figure 2.5 Validation of pigmentation candidate genes using Exelixis insertion mutants and RNAi knockdown. The y-axis in all panels is the deviation of the appropriate control line mean from the experimental line mean. Increases and decreases in the proportion of melanization are given above and below the x-axis, respectively. (A) Exelixis insertion mutants. (B) pnr-GAL4 x RNAi-UAS lines. (C) tub-GAL4 x RNAi-UAS lines. (D) ubi-GAL4 x RNAi-UAS-lines. ***: P<0.0001.
Figure 2.6 Female abdominal cuticles of Exelixis transposon insertion lines with significant effects on pigmentation. (A) Exelixis $w^{118}$ control. (B) CG10625$^{e01211}$. (C) CG7852$^{e04511}$. (D) CG9134$^{e00098}$. (E) daily$^{d01097}$. (F) Efa6$^{e03476}$. (G) Fili$^{d04573}$. (H) Glut1$^{d05758}$. (I) kay$^{e02002}$. (J) kkv$^{e06225}$. (K) klar$^{d05910}$. (L) loco$^{d09879}$. (M) Sucb$^{e01940}$. ns = not significant. ↑ and ↓ indicate significantly increased and decreased proportions of dark melanin, respectively. T5 = tergite 5 and T6 = tergite 6.
Figure 2.7 Female abdominal cuticles of with significant effects on pigmentation in lines with targeted RNAi knockdown constructs. A-J: pnr-GAL4 x UAS-RNAi genotypes. (A) VDRC KK library control (B) btn (C) CG7852 (D) Klp61F (E) ru (F) VDRC GD library control (G) Efa6 (H) klar (I) sinu. J-Q: tub-GAL4 x UAS-RNAi genotypes (J) VDRC KK library control (K) btn (L) CG10625 (M) CG9134 (N) VDRC GD library control (O) Efa6 (P) loco. Q-T: ubi-GAL4 x RNAi-UAS genotypes (Q) VDRC KK library control (R) CG1887 (S) VDRC GD control (T) klar. ns = not significant. ↑ and ↓ indicate significantly increased and decreased proportions of dark melanin, respectively.
**DISCUSSION**

The DGRP lines show substantial natural variation in female abdominal pigmentation, ranging from lines with no dark melanin to complete melanization on tergites 5 and 6. Despite being sampled from a single population, the lines span the range of pigmentation difference between the well-studied sister species *D. yakuba* and *D. santomea*. *D. santomea* is the lightest member of the *D. melanogaster* species subgroup; however, several of the DGRP lines are lighter than *D. santomea* (e.g., DGRP_441, Fig 2.1C). Utilizing the substantial genetic variation and a mapping population that is powerful to detect common variants associated with the variation, especially those with moderate to large effects, we identified a total of 155 genetic variants associated with variation in female abdominal pigmentation using GWA analyses.

We identified variants in four genes previously shown to affect adult *D. melanogaster* pigmentation: *bab1*, *t*, *e*, and *osa* [2,16,23,24]. Most of the *bab1* SNPs were associated only with the difference between T5 and T6 pigmentation, suggesting that variation in *bab1* may underlie the genetic and phenotypic correlations between the traits. Most of the *bab1* and *t* minor alleles are at moderate frequencies in the DGRP. These SNPs could be neutral or could be maintained segregating by a balance of unknown selective forces. We identified three SNPs in the CRE of *t* that were also found in the European populations studied by Bastide *et al.* [9], indicating that these SNPs are maintained in distant populations. Of note, our study is the first to associate natural variation in pigmentation with genetic variation in the transcription factor *osa*.

A majority of the variants identified in this study are in intronic or intergenic regions. Among the total regulatory elements identified, *dl* and *cad* binding sites were the most highly represented suggesting a role for these two TFs on pigmentation patterning. Over half of the genetic variants located within *bab1* were in known regulatory regions including some for *dl* and *cad* binding. We also identified a SNP upstream of *e* that is located within a *cis*-regulatory region consistent with the studies of African and European *D. melanogaster* [9,16,19]. These results implicate *cis*-regulatory evolution, which likely limits negative pleiotropic effects, as a major contributor to phenotypic variation within the DGRP population.

In addition to genes previously known to affect *Drosophila* pigmentation, we identified many novel candidate genes. We showed that 61% of the candidate genes affect the proportion of dark pigmentation on tergites 5 and 6 using mutant and RNAi knockdown experiments. These genes are known to be involved in processes such as sugar binding and transport, vesicle formation and transport, and cuticle formation. We summarize what is known from the literature about these candidate genes below and speculate about their roles in pigmentation and phenotypic evolution.

Prior to molting and eclosion, insects accumulate tyrosine-derivatives conjugated with hydrophilic molecules such as glucose, phosphate, and sulfate in the hemolymph. This keeps the reactive
pigment precursors in an inert state until the organism is ready to molt or eclose [13,25–29]. We identified two genes, triforce (tfc, CG9134) and Glut1, which may facilitate the transport of glucose or glucose conjugates from the hemolymph to the epidermal cells. tfc is a C-type lectin with a carbohydrate binding domain and Glut1 is a membrane bound glucose transporter [30].

Several of our new pigmentation genes have roles in relatively well-described developmental pathways. These include kay, dally, Fili, and ru. kay is a transcription factor in the JNK signal transduction pathway [31]. It is required for decapentaplegic (dpp) expression, wing and leg development, and the elongation of the cells in the epidermis [32]. dpp expression in the tergite corresponds to the midline pigment stripe, and ectopic expression of dpp in pupae expands this stripe [32]. Furthermore, dpp signal transduction is potentiated by dally [33]. Additionally, dpp and Epidermal growth factor receptor (Egfr) signaling work synergistically to specify the lateral tergite cell fate [32]. ru, also known as Rhomboid-3 (Rho-3), activates Egfr signaling and thus may help determine epidermal cell fate in the developing tergites; however, there are several Rho proteins capable of this activation [34]. Fili is a transmembrane protein that is involved with apoptosis in the wing imaginal disc and we speculate may facilitate proper tergite differentiation during metamorphosis [35].

Four validated candidate genes are involved with vesicle formation and transport: pinstripe (pns, CG7852, which describes the vertical stripe of pigment remaining on T6 in the RNAi knockdown), Efa6, Klp61F, and klar. pns is predicted to have Rab guanyl-nucleotide exchange factor activity, which facilitates vesicle transport by activating Rab proteins [36,37]. Rab5 works in conjunction with Megalin to remove the Yellow protein from the tanning D. melanogaster wing [38]. Efa6 is also involved with Rab signaling and vesicle mediated transport [36]. Klp61F is a kinesin motor, and klarsicht (klar) regulates microtubule-dependent vesicle transport along microtubules. Both could be involved in transporting vesicles of enzymes and/or dopamine-derivatives to and from the cuticle. Together these genes may represent components of the little-known transport mechanism for cuticle tanning in D. melanogaster.

farmer (frm, CG10625 – the not quite fully developed stripes of pigment on tergites 5 and 6 are similar to the tan lines on the arms of a farmer after much time spent in the sun), kkv, and sinu, and loco may affect cuticle development and structure. frm is a cuticle structural protein [36]. kkv is one of two chitin synthase genes in D. melanogaster [39]. sinu is a claudin required for septate junction organization, cell-cell adhesion, and proper localization of proteins involved in chitin filament assembly in D. melanogaster [40]. loco regulates G protein signaling, and Gy1 signaling is required for proper septate junction protein localization including sinu [41,42]. These proteins may help maintain the structural integrity of the adult cuticle and our study shows that when perturbed, they affect pigmentation.
Sucb and ph may affect the organism more broadly. Sucb is a succinate-CoA synthetase in the Krebs cycle [36]. It is plausible that variation in energy production due to genetic variation in key enzymes could indirectly affect variation in adult pigmentation by altering resources available for cuticle development. ph is a CD36 homologue, and dipteran CD36 family members may have roles in transport of cholesterol and steroids during ecdysone synthesis [43]. Since ecdysone is required for proper insect development and molting, disrupted transport of ecdysone precursors may explain the severe RNAi phenotype for this gene.

In summary, we provide evidence that genetic variation at a number of steps in regulatory, developmental, and transport pathways may contribute to natural variation in abdominal pigmentation. These findings exemplify the pleiotropic nature of these genes which may limit their potential as adaptive targets [44–46] [52]. Several of the mutant and RNAi knockdown lines were lethal or had strong debilitating effects providing some support of this. It is known that the large-effect genes, t and e, are pleiotropic as well [23]. The difference may be that not all pigmentation genes have the necessary regulatory modules to alleviate pleiotropic effects. However, these candidate genes may contain tissue- or stage-specific gene regulatory architectures since most of the GWAS associated SNPs are in intronic and intergenic regions. Furthermore, a distinction should be made between pleiotropic genes and pleiotropic variants. A given gene may be pleiotropic, while particular variants within that gene may not be [47]. Additionally, in the DGRP lines allelic variants are homozygous. In nature, alleles with detrimental effects may be tempered in the heterozygous state or epistatic interactions may arise with differing combinations of polymorphic loci.

Consistent with other DGRP studies, we identified many more genetic variants associated with pigmentation than previous studies [48–51]. We suspect earlier studies may have only had the ability to identify the major effect loci and missed the more polygenic variation at these other loci. Most used only a small number of fly lines and thus interrogated a smaller sample of allelic variation, analyzed only known pigmentation genes, or the limited sample size of the mapping population gave reduced statistical power to detect variants with smaller effects. For example, the Winter's Lines, a panel of 144 recombinant inbred lines used to map the effect of bab1 on pigmentation in female D. melanogaster, were generated from only two gravid wild caught females [2]. The study of Bastide et al. [9] pooled individuals with extreme phenotypes for sequencing. This approach may filter out variants that lead to intermediate phenotypes and select for large effect variants. Pool and Aquadro focused only on e sequences among the 25 African lines eliminating any possibility of identifying additional loci [16]. The DGRP is more representative of natural populations and harbors many more polymorphic loci that may contribute to phenotypic variation and evolution [52]. Given both the population sample and the genome-wide coverage of polymorphisms, this study is perhaps the most comprehensive analysis of variation in Drosophila pigmentation to date.
It is important to acknowledge that gene expression knockdown and mutant analyses are only an approximate confirmation of causative SNPs. Genes implicated by the GWA analyses that do not confirm in these functional tests may be true positives and contribute to variation in pigmentation, but they do not change pigmentation when gene expression is reduced. Future studies should test effects of individual SNPs and further characterize the mechanisms though which the candidate genes affect variation in pigmentation; their potential interactions with variants in the candidate genes with major effects; and their allele frequency distributions in different populations. These studies will help elucidate the contribution of these novel variants to adaptive phenotypic evolution or whether they are population-specific deleterious variants that are maintained segregating by mutation-selection balance.

Our results open the door for new hypotheses to be tested on the transport of dopamine derivatives and conjugates from the hemolymph to the cuticle, the formation and movement of vesicles within epidermal cells, the mechanisms of regulatory and developmental pathways during tergite differentiation, the interactions of chitin filaments with cell adhesion and cuticle proteins, and how metabolic and hormonal regulation could lead to variation in pigmentation. Genetic variants that affect these processes could potentially serve as targets of adaptive evolution or sexual selection in natural populations. This study is a start. However, much more work is needed to draw mechanistic inferences about these novel candidate genes and their contributions to the evolution of pigmentation.

**MATERIALS AND METHODS**

*Drosophila stocks and phenotyping*

The DGRP consists of 205 inbred lines with complete genome sequences. We scored female flies of 175 DGRP lines – aged 5 to 8 days – for the proportion of melanization on abdominal tergites 5 and 6. Two independent replicates for each DGRP line were reared and five individuals were scored from each replicate vial (N = 10 flies per line). The flies were reared in vials at a controlled adult density (CAD) of 10 males and 10 females on cornmeal-molasses-agar medium at 25°C, 75% relative humidity, and a 12-h light-dark cycle. The parental generation was allowed to lay eggs for 3 days. Each fly was visually assessed by a single observer for the percentage of brown/black melanin covering each tergite; the scores ranged from 0 for no dark pigmentation to 4 for 100% dark pigmentation in increments of 0.5.

*Statistical and quantitative genetic analyses*

We partitioned variation in pigmentation into genetic and environmental components using an ANOVA model of form \( Y = \mu + L + T + L\times T + R(L\times T) + \epsilon \), where \( Y \) is phenotype, \( \mu \) is the overall mean, \( L \) is the random effect of line, \( T \) is the fixed effect of tergite, \( R \) is the random effect of replicate.
vial, and $\varepsilon$ is the residual. We also performed reduced ANOVAs separately for each tergite of form $Y = \mu + L + R (L) + \varepsilon$. We estimated variance components for the random effects using REML. We computed the broad-sense heritability ($H^2$) of pigmentation for each tergite separately as $H^2 = \sigma^2_L / (\sigma^2_L + \sigma^2_\varepsilon)$, where $\sigma^2_L$ is the among-line variance component and $\sigma^2_\varepsilon$ is the error variance. We computed the genetic correlation between the tergites ($r_{T5,T6}$) as $\text{Cov}_{T5,T6} / \sigma_{LT5} \sigma_{LT6}$, where $\text{Cov}_{T5,T6}$ is the covariance in pigmentation score between tergites 5 and 6. All analyses were performed with version 9.3 of the SAS System for Windows (2013 SAS Institute Inc.).

**Genome-wide association analysis**

To identify genomic regions harboring variants contributing to natural variation in the proportion of tergite melanization, we conducted a GWA study for each tergite. The DGRP lines are also segregating for *Wolbachia* infection and for the following common inversions: $\text{In}(2L)t$, $\text{In}(2R)NS$, $\text{In}(3R)P$, $\text{In}(3R)K$, and $\text{In}(3R)Mo$. We performed GWA studies in two stages. In the first stage, we adjusted the line means for the effects of *Wolbachia* infection and major inversions. We then used the adjusted line means to fit a linear mixed model in the form of $Y = Xb + Zu + e$, where $Y$ is the adjusted phenotypic values, $X$ is the design matrix for the fixed SNP effect $b$, $Z$ is the incidence matrix for the random polygenic effect $u$, and $e$ is the residual. The vector of polygenic effects $u$ has a covariance matrix in the form of $A\sigma^2$, where $\sigma^2$ is the polygenic variance component. We fitted this linear mixed model using the FastLMM program (version 1.09) [53]. We performed these single marker analyses for the 1,897,337 biallelic variants (SNPs and indels) with minor allele frequencies $\geq 0.05$ whose Phred scale quality scores were at least 500 and genotypes whose sequencing depths were at least one and genotype quality scores at least 20 [15]. All segregating sites within lines were treated as missing data. Additionally, we performed single marker tests for association on line means that were adjusted for the effects of *Wolbachia* infection and major inversions but not corrected for the relationship matrix. Significant variants were annotated using the 5.49 Release of the FlyBase annotations.

**Variance in pigmentation explained by top variants**

For each variant, we calculated two variance components. First, to calculate the variance explained by a variant without adjusting for variants in known pigmentation genes, we fitted a linear model for the adjusted line means for only the focal variant and used the $R^2$ of the model to represent the variance explained by it. Second, to calculate additional variance explained by a focal variant after accounting for variants in known pigmentation genes, we first used stepwise selection to select the top associations for each pigmentation gene (tan, ebony, or bab1), requiring $P$-values to be smaller than $10^{-5}$ if more than one variant entered the model, and no $P$-value requirement if there was only one variant. The $R^2$ of this baseline model (different for each of the four traits) is the variance
explained by the pigmentation genes. We added each focal variant to the baseline model and calculated the difference between the $R^2$ of the new model and the $R^2$ of the baseline model, which represented the additional variance explained by the variant after accounting for the pigmentation genes. To calculate total variance explained by all significant variants, we used a mixed model approach because of the large number of variants. We computed the variance/covariance matrix based on the genotype matrix and estimated the variance components using the rrBLUP R package.

Validation of candidate genes

We tested 12 of the 13 genes implicated by the T6 pigmentation GWA analysis, none of which were previously known to affect variation in pigmentation in *D. melanogaster*: CG33298, Fili, Vmat, mwh, Klp61F, CG9134, CG7852, CG1887, klar, Glut1, Efa6, and btn. From the T5 pigmentation GWA analysis, we selected candidate genes that (1) had an Exelixis mutant [20] or VDRC RNAi [22] line available at the time of the study; (2) are involved in development, especially of the cuticle or epidermis, or pigmentation according to FlyBase and the available literature; and (3) show mRNA expression patterns similar to the regulatory genes, *bab1* and *Dsx*, and genes in the pigmentation biosynthesis pathway (such as *t* and *e*), a peak of expression at 24 hr after puparium formation and 2-4 days after puparium formation, respectively, according to the modENCODE tissue and temporal expression data [27,52]. This resulted in 16 additional candidate genes: ru, CG10625, sinu, Sucb, dally, CG32052, Nc, Cerk, kkv, CG15803, loco, TwdlC, kay, CG1340, CG42594, and CG42340. For each candidate gene, we tested either an Exelixis transposon insertion line [20], a VDRC RNAi line [54], or when possible, both a mutation and RNAi construct. We assessed the proportion of melanization for both T5 and T6 for all candidate genes.

We evaluated 15 Exelixis transposon insertion lines: CG33298$^{d10678a}$, Fili$^{d04573}$, mwh$^{d01620}$, Klp61F$^{d02870}$, CG9134$^{e00088}$, CG7852$^{d04511}$, klar$^{d05910}$, Glut1$^{d05758}$, Efa6$^{d03476}$, CG10625$^{d01211}$, Sucb$^{d01940}$, dally$^{d01097}$, kkv$^{c06225}$, loco$^{d09870}$, and kay$^{d02002}$. The Exelixis progenitor w$^{118}$ line was used as a control. The KK and GD library progenitor lines were used to make control crosses for the RNAi knockdown experiments. Males from the GAL4 driver line were crossed to virgin females of the VDRC UAS line for all crosses. Three GAL4 driver lines were used for the RNAi crosses. All VDRC UAS lines were crossed with the full-body tubulin-GAL4/Sb driver and a pannier-GAL4 driver ($y^1$ w$^{118}$; P(w$^{mW,hS}=GawB)pn$M0237/TM3, P(w$^{mC}=UAS-y.C)MC2, Ser$^1$). In instances of lethality with the tubulin-GAL4/Sb driver, the lines were crossed to another full-body ubiquitin-GAL4/Cy driver. All GAL4-driver lines were obtained from the Bloomington, Indiana *Drosophila* Stock Center. We tested in total 26 RNAi knockdown constructs: CG33298, Fili, Vmat, mwh, Klp61F, CG9134, CG7852, CG1887, klar, Efa6, btn, ru, CG10625, sinu, dally, CG32052, Nc, Cerk, kkv, CG15803, loco, TwdlC, kay, CG1340, CG42594, and CG42340.
We reared three independent replicates for each Exelixis transposon insertion line, for each RNAi cross and for the appropriate controls under the same conditions as the DGRP lines, but in 8 oz. bottles with a controlled adult density of 20 males and 20 virgin females. We scored the proportion of melanization on T5 and T6 for 50 5-8 day old female progeny per replicate (N = 150 flies per genotype) from each Exelixis line or RNAi cross. In a few instances where viability was low fewer than 50 individuals per replicate were scored: \textit{pnr-GAL4 x sinu} (N = 23), \textit{ubi-GAL4 x CG1887} (N = 90), and \textit{ubi-GAL4 x klar} (N = 95). Means of test lines were compared to those of the appropriate controls with a Dunnett’s test, which corrects for multiple testing, using JMP Pro 10.0.0 (2012 SAS Institute Inc.)

Dissection and photography
After mutant lines and RNAi knockdown progeny were scored for pigmentation, they were preserved in a 3:1 ethanol/glycerol solution and stored at 4°C until dissection for imaging. The fly cuticles were dissected from the abdomen and mounted to a glass slide using Permount and a glass cover slip. All photographs were taken with an Olympus DP25 microscope camera on an Olympus SZ61 stereo microscope.

ACKNOWLEDGMENTS
We thank Aki Yamamoto, Gunjan Arya, and Lavanya Turlapati for ordering \textit{Drosophila} stocks and supplies. We acknowledge the work of Alden Hearn, Thomas Freeman, and Lenovia McCoy who prepared the cornmeal-molasses-agar medium each week and helped with laboratory up-keep. We also thank Robert Anholt for helpful discussions and comments on the manuscript. We acknowledge the Vienna Drosophila RNAi Center, the Exelixis Collection at the Harvard Medical School, and Bloomington Drosophila Stock Center. We thank the three anonymous reviewers and the journal editors for PLoS Genetics for providing constructive criticism and feedback that greatly improved the manuscript.

REFERENCES


CHAPTER 3. THE GENETIC ARCHITECTURE OF NATURAL VARIATION IN CUTICLE COLOR OF DROSOPHILA MELANOGASTER

Lauren M. Dembeck\textsuperscript{1,2}, Rachel Lyman\textsuperscript{1,2,3}, Wen Huang\textsuperscript{1,2}, Akhila Boyina\textsuperscript{4}, Hannah Moyles\textsuperscript{5}, and Trudy F.C. Mackay\textsuperscript{1,2}
\textsuperscript{1}Department of Biological Sciences, Program in Genetics and \textsuperscript{2}W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695, USA
\textsuperscript{3}Division of Biology and Biomedical Sciences, Evolution, Ecology and Population Biology Program, Washington University, St. Louis, Missouri 63110, USA
\textsuperscript{4}William G. Enloe High School, Raleigh, NC 27610, USA
\textsuperscript{5}Wake Forest-Rolesville High School, Wake Forest, NC 27587, USA

ABSTRACT
Body coloration is an ubiquitous and ecologically important trait across animal taxa. In the fruit fly, Drosophila melanogaster, the cuticle-coloring melanin-sclerotin biosynthetic pathway has been well-studied; however, the mechanism by which the pigment precursors are moved from the epidermal cell to the overlying cuticle has not been described. To identify candidate genes in this transport pathway, we characterized natural variation in female abdominal coloration in the D. melanogaster Genetic Reference Panel and conducted genome-wide association analyses. We identified 72 naturally occurring genetic variants in or near 58 genes. We use mutant and RNAi knockdown analyses to show that four genes, mgl, kirre, Trim9, and spz5, affect body coloration. These results provide evidence that genes involved in vesicle dynamics contribute to variation in body coloration and suggest testable hypotheses of the mechanistic basis of epidermal-cuticle dynamics.

INTRODUCTION
Body coloration is an ecologically and behaviorally important trait across animal taxa. It provides information to conspecifics about species and sex identity, mating status, and health [1]. It also protects organisms from abiotic factors such as ultraviolet radiation and radioactivity [2]. Genes involved in color patterning tend to be highly pleiotropic. For example, animal domestication is often accompanied by phenotypic changes such as floppy ears, curly tails, piebald spotting and coat mottling [3,4]. Aggression in cichlids is predictable based on color differences [5]. Drosophila melanogaster tan (t), yellow (y), pale (ple), and ebony (e) mutants are all known to have behavioral differences in addition changes in body coloration.
The biosynthesis pathway for pigmentation has been well studied in *D. melanogaster*, but there is still little known about how the dopamine-derivatives that underlie coloration are transported from the epidermal cells to the cuticle [7-9]. Here, we used the *D. melanogaster* Genetic Reference Panel (DGRP) of sequenced, inbred lines to perform a genome wide association study (GWAS) to identify genetic variants contributing to natural variation in the intensity of melanin and sclerotin. None of the major effect genes previously associated with variation in the proportion of melanin covering the abdominal tergites (t, bab1 or e) were associated with these color intensity traits [7–9]. However, we identified variants in 53 other genes associated with variation in color intensity including a gene known to be involved in wing cuticle development, *Megalin* (*mgl*). We tested mutant and/or RNAi-knock down alleles of 12 candidate genes. In addition to confirming the need for *mgl* for proper melanin deposition, we demonstrate that three genes that are involved in cell adhesion and vesicular transport, *spatzle 5* (*spz5*), *kin of irre* (*kirre*), and *Tripartite motif containing 9* (*Trim9*, are also involved in properly coloring the tergites of *D. melanogaster* females.

**RESULTS**

*Quantitative genetics of coloration*

We measured the grayscale color intensity of melanin and sclerotin on tergites 2 – 7 for 162 DGRP lines. The ten-point average melanin or sclerotin intensity scores were highly correlated among the tergites. For sclerotin, T2-5 are the most highly correlated ($r = 0.87 – 0.98$). These tergites have even stronger correlations for melanin ($r = 0.94 – 0.98$). The T6 and T7 scores are the least correlated with all other tergites, both within and between the traits: sclerotin ($r = 0.60 – 0.83$) and melanin ($r = 0.43 – 0.89$). This result reflects the modular nature of abdominal pigmentation in *Drosophila*. Although the correlation among most tergites is high for both traits, some lines have little to no light sclerotin on tergites 5 and 6. Therefore, we chose individual tergites for further analysis of each color trait rather than averaging the scores for each trait across all tergites. We used the ten-point average data from tergite 3 for sclerotin intensity and tergite 6 for melanin intensity (Fig 3.1).

Averaged across all lines, the mean T6 melanin and T3 sclerotin scores were 125 and 196, respectively. There was significant genetic variation in intensity among lines for both melanin ($P_M = 4.36 \times 10^{-183}$) and sclerotin ($P_S = 1.52 \times 10^{-91}$), with broad sense heritabilities ($H^2$) of $H^2_M = 0.81$ and $H^2_S = 0.62$, respectively. The phenotypic ($r_{P(M,S)} = 0.63 \pm 0.06$) and genetic ($r_{G(M,S)} = 0.68 \pm 0.05$) correlations ($\pm$ standard error) between the intensity for melanin and sclerotin were moderate but different from unity, suggesting they have partially different genetic bases. The high broad sense heritabilities for these abdominal coloration traits provided a favorable scenario for GWA analyses.
 Genome-wide association analyses

We performed GWAS on the T6 melanin and T3 sclerotin traits to identify variants associated with female abdominal coloration traits. In addition to the individual trait GWAS, we also performed these analyses for the difference of the two coloration traits.

The DGRP lines vary in Wolbachia infection status and karyotype for several common polymorphic inversions. We did not find significant associations of Wolbachia infection ($P_M = 0.23$, $P_A = 0.12$, and $P_D = 0.46$) nor inversion karyotype on the average intensity of melanin or the average, or difference of the two color traits. However, the average intensity of sclerotin was significantly affected by Wolbachia infection status ($P_S = 0.008$) and $\ln(3R)K$ ($P_S = 0.04$). For each GWAS, we used a mixed model that accounted for any effects of Wolbachia, inversions, and cryptic relatedness [10].

Combining all of the GWAS models, we identified 72 variants associated with coloration at a nominal reporting threshold of $P < 10^{-5}$. Although most variants do not reach individual Bonferroni-corrected significance levels, quantile-quantile plots indicate a systematic departure from random expectation below $P < 10^{-5}$, justifying our choice of this reporting threshold and suggesting that the top associations are enriched for true positives. Of these, 30 were associated with melanin intensity, 23 with sclerotin intensity, and 37 with the difference in pigmentation between melanin and sclerotin intensity. A total of 58 candidate genes were implicated by these associated variants. One SNP ($2R_{17958592}$_SNP), exceeded a strict Bonferroni correction for multiple tests ($P = 2.61 \times 10^{-5}$). It is the most highly associated SNP with the average intensity of the two color traits ($P_A = 1.12 \times 10^{-5}$) and well as for melanin intensity ($P_M = 4.75 \times 10^{-5}$), but is also highly associated in the sclerotin ($P_S = 6.00 \times 10^{-7}$) and difference ($P_D = 7.84 \times 10^{-7}$) GWAS. In all cases, the effect of this SNP estimated as the average difference in color between lines with the major and minor alleles, was negative, i.e., the minor allele was associated with a darker cuticle than the major allele. The SNP causes a synonymous change in the gene $CG11475$ and is located 147 bp downstream of $CG11474$.

Among the genes harboring variants associated with variation in melanin intensity and the difference between melanin and sclerotin intensity we find two SNPs in a gene that regulates vesicle endocytosis during melanization of wing cuticle in $D$. melanogaster, Megalin (mgl), as well as an additional large group of novel candidate genes. A SNP in sponge (spg, $3R_{24693666}$_SNP), a Rho GTPase involved in cytoskeletal dynamics including lamellipodium formation, was the second most strongly associated in these three GWAS ($P_M = 8.52 \times 10^{-5}$, $P_A = 8.98 \times 10^{-5}$, and $P_D = 2.36 \times 10^{-7}$).

Variants associated with sclerotin intensity have both negative and positive effects. In addition to the previously mentioned Bonferroni-significant SNP was a SNP in Tripartite motif containing 9 (Trim9, $2L_{10616448}$_SNP, $P_S = 1.33 \times 10^{-6}$) and a SNP in another gene implicated in synaptic vesicle endocytosis, Bicaudal D (BicD, $2L_{17460458}$_SNP, $P_S = 8.18 \times 10^{-6}$).
Sixteen variants were associated with only the difference in intensity between the two color traits; nearly half of these were located in intergenic regions. Three were located within an intron of mgl (X_9300857_SNP, \( P_D = 1.37 \times 10^{-6} \); X_9300887_SNP, \( P_D = 4.48 \times 10^{-6} \); and X_9300675_SNP, \( P_D = 5.05 \times 10^{-6} \)). SNPs in the immune regulator spatzle 5 (spz5, 3L_2884825, \( P_D = 4.71 \times 10^{-6} \)) and Rap GTPase activating protein 1 (rapgap1, 2L_7509428_SNP, \( P_D = 3.42 \times 10^{-6} \)) were also strongly associated.

A majority of variants associated with sclerotin intensity (21/23) were associated only with that trait. In contrast, all of the variants associated with melanin intensity were also associated in the GWA analyses for the difference of the traits. In agreement with the low genetic correlation, this indicates that the two traits are largely independent and that associations with the difference are largely driven by variation in melanin intensity, which varies more than sclerotin intensity (Fig 3.1).

**Functional validation studies**

We focused our subsequent analyses on six candidate genes with known roles in vesicle formation or transport and cell adhesion: spz5, tutl, dpr6, CG42346, kirre, and Trim9; the two computationally predicted genes implicated by the SNP which exceeds the Bonferroni correction, CG11474 and CG11475; mgl; and three other genes involved with vesicle and cytoskeletal dynamics, BicD, and rapgap1, and spg.

We attempted to test the effects of each candidate gene on T3 sclerotin and T6 melanin coloration using both a mutant from the Exelixis Collection at the Harvard Medical School and Vienna Drosophila RNAi Center (VDRC) RNAi knockdown line when available. Of the 12 RNAi knock down alleles, five (tutl, CG42346, rapgap1, BicD, CG11475) were lethal when ubiquitously expressed. Mutations were not available for three of these candidate genes, and the tutl mutant was homozygous lethal. Thus we were not able to measure the effects of gene or expression alterations on pigmentation for tutl, CG42346, CG11475, BicD. We were able to test dpr6, kirre, spz5 and Trim9 for both mutant and RNAi knock down alleles; rapgap1 and spg for a mutant allele; and mgl and CG11474 for RNAi constructs.

Of the mutant lines tested, spz5 lightened sclerotin color (\( P = 0.01 \), Fig 4.3.2 A, Fig 4.3.3). No other mutant alleles of candidate genes affected T3 sclerotin or T6 melanin color intensity (Fig 3.2 A, C). In the RNAi knockdown experiments, melanin and sclerotin color were altered for four genes, mgl, kirre, spz5, and Trim9 (\( P = 0.02 \) for mgl and \( P < 0.0001 \) for all; Fig 3.2, Fig 3.4). All showed darkening of both T3 sclerotin and T6 melanin relative to the control (Fig 3.2B, D).
Figure 3.1 Natural variation in T3 sclerotin and T6 melanin coloration in the DGRP. (A) T3 sclerotin (tan) and T6 melanin (black). DGRP lines are in order from darkest to lightest melanin on T6. (B) Images display one half of the fly abdominal cuticle, split along the dorsal midline. DGRP lines are denoted by DGRP-XXX below each image.
Figure 3.2 Validation of coloration candidate genes using mutants and RNAi knockdown. The y-axis in all panels is the deviation of the appropriate control line mean from the experimental line mean. Lightening or darkening of the traits is given above and below the x-axis, respectively. (A) Exelixis Collection mutants T3 sclerotin and (C) T6 melanin scores. (B) ubiquitin-GAL4 x RNAi-UAS-lines T3 sclerotin and (D) T6 melanin scores. *: $P < 0.05$, **: $P < 0.01$ ***: $P < 0.0001$. 
Figure 3.3 Representative images of mutant female abdominal cuticles with significant effects on T3 coloration. Each row exhibits a representative image for each line followed by a magnified portion of the T3 sclerotin in full-color and the 8-bit grayscale version. The average score for each line is denoted on the left.
Figure 3.4 Portions of T6 melanin and T3 sclerotin from lines expressing targeted RNAi knockdown constructs with effects on coloration. Each row exhibits the full-color cuticle sample followed by the 8-bit grayscale version from a given line for T6 (left) and T3 (right). The center column image is the line's representative image from which the cuticle samples were taken. The average score for each line for each trait is denoted on the left. All lines are from the VDRC KK and crossed to *ubi-GAL4*. *ns: not significant.
**DISCUSSION**

Most *Drosophila* pigmentation studies to date have measured variation in the proportion, or presence or absence of melanin on the abdomen [10–12,27–29] (though see the custom spectrometry system of [30] and the combinatorial approach in [11]). Accordingly, these studies have associated genetic variants in genes responsible for spatial patterning and melanin production such as *bab1*, *t*, and *e*. We used an alternative phenotyping strategy to characterize natural variation in female abdominal tergite coloration in the DGRP. Using genome-wide association analyses in conjunction with mutant and RNAi functional validation experiments, we identified three genes, *kirre*, *Trim9*, and *spz5*, that alter body color intensity, and confirmed the effect of reduced *mgl* expression on melanin deposition. Previous studies indicate that these genes have roles in cell adhesion and vesicle transport. These roles are consistent with a possible mechanism for epithelium-to-cuticle transport of pigmentation enzymes and dopamine-derivatives.

*mgl* mediates the endocytic removal of the Yellow protein, which is required to produce black melanin, from the developing cuticle. When *mgl* expression is reduced, Yellow accumulates leading to the ectopic formation of abnormally large melanin granules [15]. Our finding of darkened abdominal melanin in *mgl*-RNAi progeny is consistent with these results, and our GWA analysis is the first to associate natural genetic variation in *mgl* with variation in abdominal coloration in *Drosophila*.

*Trim9*, *kirre*, and *spz5* have pleiotropic effects including roles in neuronal function, cell adhesion, and dorsal-ventral body patterning. Trim proteins are found in all metazoans and have many intracellular functions. *Trim9* binds to microtubules and regulates synaptic vesicle exocytosis [16]. Similarly, *kirre* and its orthologs in other model organisms have been linked to renal and neuronal activity [17–19] and sensitivity of muscle development to methylmercury exposure [20]. Mutations in the vertebrate homologue, *Neph1*, exhibit congenital nephrotic syndrome in humans and mice [17,18], and *kirre* mutants recently have been developed as models for kidney disease in *D. melanogaster* [19,20]. The *C. elegans* ortholog, *syg-1*, is involved in synapse formation and function; in the *syg1* ky652 mutant, synaptic vesicles fail to move properly and accumulate forming ectopic clusters [21]. Cross-species replacement of *Neph1* in *C. elegans syg1* mutants can restore neuronal functioning [22]. These findings are consistent with those of *D. melanogaster* pigmentation mutants and their effects on neuronal function and thus behavior.

Both mammalian and invertebrate Toll genes are required for innate immunity. *spz5* is a ligand for the Toll 6 and Toll 7 receptor proteins [23]. Overexpression of Toll-related proteins including Toll 6 and Toll 7 in *D. melanogaster* leads to developmental abnormalities and lethality [24]. In *Aedes aegypti*, the *spz5* ortholog is up-regulated in hemocytes, cells critical for immune mechanisms, including melanization, of mosquitoes infected with pathogenic bacteria [25].
Our results, in conjunction with previous studies of these candidate genes, support the idea of extensive pleiotropy through a basal cellular mechanism. Coloration, cell adhesion, and renal and neuronal function all require vesicle formation and dynamics. We hypothesize that the genes we have identified may be part of a fundamental system for vesicle dynamics giving rise to this extensive pleiotropy in D. melanogaster.

MATERIALS AND METHODS

**Drosophila stocks and phenotyping**

We scored female flies of 162 DGRP lines for the intensity of melanin and sclerotin on abdominal tergites 2 - 6. Each DGRP line was reared at a controlled adult density of 10 males and 10 females on cornmeal-molasses-agar medium at 25°C, 75% relative humidity, and a 12-h light-dark cycle. The parental generation was allowed to lay eggs for 3 days. Upon eclosion, virgin females were collected and aged for four days under the same conditions. On the fourth day, the flies were placed in microcentrifuge tubes containing a 3:1 ethanol: glycerol storage solution. This solution preserves the flies without allowing them to become desiccated or faded [14]. The flies were then stored at 4 °C until dissection.

**Abdominal cuticle dissection and imaging**

The abdominal cuticle was dissected from five female flies of each DGRP line; for nine lines fewer than five individuals were used due to viability issues or developmental abnormalities such as missing tergites. Each cuticle was then mounted on a glass slide with Permount under a glass cover slip. All fly cuticles were imaged on a white background with consistent ring lighting, magnification (4.5x), and exposure time (33 ms) using an Olympus DP2-BSW camera in a single session.

Image J [26] was used to collect 8-bit grayscale intensity data ranging from 0 (black) to 255 (white). Ten points were selected on tergites 2 - 6 of each fly cuticle image within the stripe of dark melanin and in a separate analysis, in the yellowish-tan sclerotin.

**Statistical and quantitative genetic analyses**

We partitioned variation in coloration into genetic and environmental components using an ANOVA model of form $Y = \mu + L + \varepsilon$, where $Y$ is phenotype, $\mu$ is the overall mean, $L$ is the random effect of line, and $\varepsilon$ is the residual. Variance components were estimated for the random effects using REML. We computed the broad-sense heritability ($H^2$) of coloration for each trait as $H^2 = \sigma^2_L / (\sigma^2_L + \sigma^2_\varepsilon)$, where $\sigma^2_L$ is the among-line variance component and $\sigma^2_\varepsilon$ is the error variance. We computed the genetic correlation between the regions $(r_{S,M})$ as $\text{Cov}_{S,M} / \sigma_S \sigma_M$, where $\text{Cov}_{S,M}$ is the covariance in
intensity scores between the sclerotin and melanin. All analyses were performed with version 9.3 of the SAS System for Windows (© 2013 SAS Institute Inc.).

**Genome-wide association analysis**

We performed a GWA study to identify genomic regions harboring variants contributing to natural variation in the sclerotin and melanin intensity. The DGRP lines are segregating for *Wolbachia* infection and for the following common inversions: *In(2L)t*, *In(2R)NS*, *In(3R)P*, *In(3R)K*, and *In(3R)Mo*. We performed GWA studies in two stages. In the first stage, we adjusted the line means for the effects of *Wolbachia* infection and major inversions. We then used the adjusted line means to fit a linear mixed model in the form of $Y = Xb + Zu + e$, where $Y$ is the adjusted phenotypic values, $X$ is the design matrix for the fixed SNP effect $b$, $Z$ is the incidence matrix for the random polygenic effect $u$, and $e$ is the residual. The vector of polygenic effects $u$ has a covariance matrix in the form of $A\sigma^2$, where $\sigma^2$ is the polygenic variance component. We fitted this linear mixed model using the FastLMM program (version 1.09) [27]. We performed these single marker analyses for the 1,915,397 biallelic variants (SNPs and indels) with minor allele frequencies $\geq$ 0.05 whose Phred scale quality scores were at least 500 and genotypes whose sequencing depths were at least one and genotype quality scores at least 20 [10]. All segregating sites within lines were treated as missing data.

**Candidate gene validation experiments**

We conducted RNAi knockdown experiments and used mutants to test if perturbing or disrupting candidate gene expression affects coloration. We focused on 12 candidate genes: *spz5*, *tutl*, *dpr6*, *CG42346*, *kirre*, *Trim9*, *CG11474*, *CG11475*, *spg*, *mgl*, *BicD*, and *rapgap1*. We crossed each *UAS-RNAi* line and the VDRC KK library progenitor control with a whole-body ubiquitin-GL4 driver line (*Ubi-GAL4/CyO*). We used the Exelixis Collection progenitor *w^1118* line as a control for the mutant experiment. All lines and RNAi progeny were reared under the same conditions as the DGRP lines.Virgin females were collected and aged for four days. Then ten flies were dissected, mounted, and imaged as previously described. The sample size of the *Ubi-GAL4* x *UAS-mgl* RNAi analysis was $n = 3$ due to high mortality by the end of day four. Experimental line means were compared to control line means using a Dunnett's test in JMP version 10 (© 2013 SAS Institute Inc).

**AUTHORS’ CONTRIBUTIONS**

LMD reared, dissected, and imaged all flies. LMD, RL, AB, and HM participated in data analysis; LMD and WH carried out the statistical analyses; LMD and TFCM conceived of the study, designed the study, coordinated the study, and wrote the manuscript. All authors gave final approval for publication.
ACKNOWLEDGEMENTS

We thank Aki Yamamoto and Richard Lyman for providing dissecting tools and assistance in microscopy and imaging. We thank Aki Yamamoto, Richard Lyman, Gunjan Arya, and Lavanya Turlapati for ordering Drosophila stocks and supplies. We acknowledge the work of Alden Hearn, Thomas Freeman, Lenovia McCoy and Faye Lawrence who prepared the cornmeal-molasses-agar medium each week and helped with laboratory up-keep. We also thank Robert Anholt for helpful discussions and comments on the manuscript. We acknowledge the Vienna Drosophila RNAi Center, the Exelixis Collection at the Harvard Medical School, and Bloomington Drosophila Stock Center.

REFERENCES


CHAPTER 4. THE GENETIC ARCHITECTURE OF NATURAL VARIATION IN CUTICULAR HYDROCARBON COMPOSITION IN DROSOPHILA MELANOGASTER

Lauren M. Dembeck¹,²,³, Katalin Böröczky²,³,⁴,* Wen Huang¹,²,³, Coby Schal²,³,⁴, Robert R. H. Anholt¹,²,³, and Trudy F. C. Mackay¹,²,³,⁺

¹ Department of Biological Sciences ² Genetics Program ³ W. M. Keck Center for Behavioral Biology and ⁴ Department of Entomology, North Carolina State University, Raleigh, NC 27695, USA
*Current address: Department of Neurobiology and Behavior and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA
⁺Corresponding Author: trudy_mackay@ncsu.edu, 1-919-515-5810

Keywords: Genome-wide association study; Drosophila melanogaster Genetic Reference Panel; cuticular lipids; African Drosophila; multivariate analysis

ABSTRACT

Insects coat their epicuticle with fatty-acid derived lipids that prevent desiccation (mainly cuticular hydrocarbons, CHCs). Some of these CHCs have been co-opted to serve as chemical signals that mediate social interactions and may serve as targets for adaptive evolution. While the composition of CHCs for Drosophila melanogaster has been studied extensively, little information is available about the genetic basis that accounts for individual variation in CHC composition. Here, we quantified natural variation in relative abundances of CHC components in the D. melanogaster Genetic Reference Panel (DGRP) lines. Because of the multicolinearity among the CHCs, we used principal component (PC) analysis to extract leading PCs that explained the majority of CHC variation. We identified polymorphisms in or near 305 and 173 genes in females and males, respectively associated with variation in these PCs. In addition, 18 DGRP lines contain the functional Desat2 allele characteristic of African and Caribbean D. melanogaster females (more 5,9-heptacosadiene and less of the primary female sex pheromone, 7,11-heptacosadiene). Disruption of expression of 24 candidate genes showed significant effects on CHC composition in at least one sex. These genes are associated with different stages of fatty acid metabolism and, thus, represent mechanistic targets that can give rise to individual variation in CHC composition.
“The variety of chain lengths and the number and positions of methyl branches and double bonds provide the insect with the chemical equivalent of the variable colored plumage of birds.” – Blomquist and Bagnères (1)

INTRODUCTION

The evolution of mechanisms for desiccation resistance was critical for insects to colonize dry land. A major innovation that insects evolved was the ability to produce and accumulate species-specific blends of fatty acid-derived apolar lipids over the epicuticle; the most prominent of these are cuticular hydrocarbons (CHCs) (2). CHCs are produced continuously by oenocytes, and are transported through the hemolymph and then to the cuticular surface through specialized pore canals (1,3,4). The primary role of CHCs is desiccation resistance (5,6), but they have been co-opted to serve as chemical signals mediating social interactions (7–9). These interactions include species and nest mate recognition, assessment of reproductive status, and mate choice (1,10). Thus, understanding the genetic basis for individual variation in CHC composition will shed light on evolutionary mechanisms of assortative mating (11–14) and evolution of social organization (15,16), and may facilitate the development of semiochemical-based pest control methods (17).

The extensive genetic resources available for Drosophila melanogaster make it a valuable model for studying the genetic basis of CHC production and natural variation in CHC composition. Mature D. melanogaster have sexually dimorphic CHCs ranging from chain lengths of 21 to 31 carbons (C21 – C31) (18,19). Males produce predominantly shorter-chained CHCs (< C26) and use two of the CHCs, 7-tricosene (7-C23:1) and 7-pentacosene (7-C25:1), as sex pheromones. Females produce predominantly longer-chained dienes, among which 7,11-heptacosadiene (7,11-C27:2) and 7,11-nonacosadiene (7,11-C29:2) act as the primary female sex pheromones (19–21).

The Desat locus was the first desaturase gene sequence described in insects (22) and has been implicated in the biosynthesis of pheromones. It consists of two desaturase genes, Desat1 and Desat2. Desat1 is expressed in both sexes and encodes a Δ-9 desaturase that catalyzes the synthesis of palmitoleic acid, an ω-7 fatty acid and precursor to 7-monoene and the first double bond of 7,11-diene hydrocarbons (23,24). Desat2 has a female-specific effect on CHC production and has been associated with adaptive divergence of African and Cosmopolitan races of D. melanogaster ((25) but see (26)). Desat2 encodes a functional desaturase in African females but is inactive in Cosmopolitan females due to a 16-bp deletion in the promoter region (27–29). Females with an intact Desat2 gene produce altered CHC profiles which are high in the pheromonal CHC isomers, 5,9-C27:2 and 5,9-C29:2, and low in the 7,11-isomers. African D. melanogaster females also exhibit a strong behavioral bias against non-African males (30–32).
Other genes encoding female-specific or female-biased CHCs include *DesatF* and *elongase F (eloF)* (22). *DesatF* is responsible for formation of the second double bond in the 5,9- and 7,11-dienes of females (33,34). *eloF* RNAi knockdown females display an increase in 7,11-C25:2 and a corresponding decrease in 7,11-C29:2. Thus, *eloF* appears to be involved in the production of long-chain CHCs (35).

Additionally, the gene encoding the oxidative decarbonylase that converts long-chained aldehydes to hydrocarbons in both sexes was identified as *Cyp4G1*, which encodes a cytochrome P450 (36). There are however many other enzymatic reactions necessary to produce CHCs from acetyl-CoA, malonyl-CoA, and in the case of methyl-branched CHCs, amino acid-derived precursors. CHC biosynthesis starts with fatty acid production. The fatty acid synthase gene, *FASN1*, is known but other genes involved in fatty acid synthesis and CHC conversion and turnover have not been identified.

A comprehensive analysis of CHCs in both sexes segregating in a panel of recombinant inbred lines derived from a natural population identified 25 quantitative trait loci (QTL) in females and 15 in males contributing to variation in CHCs (37), but this study did not have the power to resolve QTLs to individual genes. QTL mapping analyses have also identified genomic regions called *small monoene quantities* (*smoq*) and *seven pentacosene* (*sept*), respectively associated with variation in the proportions of 7-C23:1 and 7-C25:1 in males (38). In a mutagenesis study, *nerd* drastically reduced 7-C23:1 production in males and altered courtship behavior (39). However, none of these male loci have been resolved to specific genes. Apart from the aforementioned genes, much remains unknown regarding the genetic basis of CHC production.

Here, we used the sequenced, inbred lines of the *D. melanogaster* Genetic Reference Panel (DGRP) (40,41) to perform genome wide association (GWA) analyses for nearly all detectable CHCs in both sexes in a scenario where all common genetic variants are genotyped and local linkage disequilibrium (LD) is sufficiently low to identify candidate genes and causal polymorphisms. We found considerable heritable genetic variation in a majority of male and female CHCs, distilled the axes of genetic variation into several principal components (PCs), and performed GWA analyses on each PC. We identified 24 candidate genes plausibly associated with CHC biosynthesis and for all of them, disruption of their expression altered CHC profiles in males, females, or both sexes. Surprisingly, we also found that the DGRP lines are segregating for the ancestral and deletion alleles in *Desat2* and some of these lines exhibit the African phenotype.
**RESULTS**

*CHCs in the DGRP*

We identified and quantified 71 female CHCs and 42 male CHCs in 169 and 157 DGRP lines, respectively (Table 4.1, Fig 4.1, S1 Table). Sixteen of these CHCs have not been previously described in *D. melanogaster*. Eight of the novel compounds were methyl-branched CHCs, seven were dienes, and one was a monoene. Nine of these compounds were detected only in females.

![Diagram of chromatograms from the DGRP](image)

**Figure 4.1** Representative male and female chromatograms from the DGRP. Male cuticular lipids of DGRP_38 are shown on the top (blue) and female CHCs of DGRP_786 are mirrored below (red). All peaks for both sexes were assigned a unique number based on its corresponding compound determined by GC-MS; thus compounds shared between the sexes carry the same number. See Table 1 for the list of compound names. Compounds not previously described in *D. melanogaster* are shown in bold typeface. Some CHC isomers were not resolved by conventional GC, so a few chromatogram peaks contain more than one CHC. pA = picoAmperes, c = cis-vaccenyl acetate, * = contaminants from CHC extraction, IS = internal standard (n-C32).
We assessed the extent to which the CHCs were genetically variable in the DGRP. All but three female CHCs and one male CHC (female peaks 26, 39, and 52 and male peak 59) showed significant among-line variation in a univariate ANOVA (S2 Table). Broad sense heritabilities in females ranged from 0.98 for 7,11-C25:2 (peak 24) to 0.22 for 6-C25:1 (peak 30). Broad sense heritabilities for males ranged from 0.97 for 7-C25:1 (peak 29) to 0 for 9-C29:1 (peak 58; S2 Table).

**African female CHC phenotypes**

Surprisingly, females of 15 DGRP lines expressed the African phenotype; *i.e.*, they have high levels of 5,9-C27:2 and low levels of 7,11-C27:2 (Fig 4.2A). Since Desat2 (cytological position 87B10) is located within the common African inversion In3R(K) (computed breakpoints: 86F1-86F11;96F11-96F14 [42]) we expected the Desat2 allele status (ancestral African or a Cosmopolitan 16-bp deletion in the promoter), CHC phenotype and inversion status to perfectly co-segregate in the DGRP lines. However, this was not the case (Fig 4.2C). A total of 17 DGRP lines with female CHC phenotype data contained the ancestral, functional Desat2 allele (Table 4.2, Fig 4.2B). There was significant variation in the 7,11- and 5,9-C27:2 peaks among these lines (5,9-C27:2 & 9-C27:1 F = 35.17, P < 0.0001; 7,11-C27:2 & 2-Me-C26 F = 16.09, P < 0.0001; S3 Table). We obtained female CHC data for two African lines, Z53 and Z30, for comparison. Females had significantly less 7,11-C27:2 (2-Me-C26 co-elutes with 7,11-C27:2) according to deletion status and correspondingly more 5,9-C27:2 (9-C27:1 co-elutes with 5,9-C27:2) (Fig 4.2C). Females from DGRP lines that were either homozygous or heterozygous for the ancestral Desat2 sequence had intermediate amounts of the 7,11- and 5,9-C27:2 CHCs compared to the DGRP lines homozygous for the deletion or the African lines, respectively.
Table 4.1 Cuticular lipids identified by GC-MS in DGRP males and females. NI = not identified; nd = not detected; bold typeface = not previously identified in D. melanogaster.

<table>
<thead>
<tr>
<th>#</th>
<th>Cuticular component</th>
<th>Retention index</th>
<th>#</th>
<th>Cuticular component</th>
<th>Retention index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>♀</td>
<td>♂</td>
<td></td>
<td>♀</td>
</tr>
<tr>
<td>1</td>
<td>heneicosane</td>
<td>2100</td>
<td>2100</td>
<td>33</td>
<td>NI</td>
</tr>
<tr>
<td>2</td>
<td>x-docosene (quantified only in ♂)</td>
<td>2179</td>
<td>2179</td>
<td>34</td>
<td>NI</td>
</tr>
<tr>
<td>3</td>
<td>x-docosene</td>
<td>nd</td>
<td>2184</td>
<td>35</td>
<td>13- &amp; 11-methylpentacosane</td>
</tr>
<tr>
<td>4</td>
<td>cis-vaccenyl acetate</td>
<td>nd</td>
<td>2189</td>
<td>36</td>
<td>5-methylpentacosane</td>
</tr>
<tr>
<td>5</td>
<td>docosane</td>
<td>2200</td>
<td>2200</td>
<td>37</td>
<td>8,12-hexacosadiene</td>
</tr>
<tr>
<td>6</td>
<td>7,11-tricosadiene</td>
<td>2259</td>
<td>nd</td>
<td>38</td>
<td>7,11-hexacosadiene</td>
</tr>
<tr>
<td>7</td>
<td>2-methyldocosane</td>
<td>2263</td>
<td>2263</td>
<td>39</td>
<td>2-methylpentacosane</td>
</tr>
<tr>
<td>8</td>
<td>9-tricosene</td>
<td>2273</td>
<td>2273</td>
<td>40</td>
<td>6,10-hexacosadiene</td>
</tr>
<tr>
<td>9</td>
<td>7-tricosene</td>
<td>2280</td>
<td>2283</td>
<td>41</td>
<td>3-methylpentacosane (or only 3-methylpentacosane)</td>
</tr>
<tr>
<td>10</td>
<td>6-tricosene</td>
<td>2285</td>
<td>2286</td>
<td>42</td>
<td>7-hexacosene</td>
</tr>
<tr>
<td>11</td>
<td>5-tricosene</td>
<td>2291</td>
<td>2291</td>
<td>43</td>
<td>6-hexacosene + impurity (i)</td>
</tr>
<tr>
<td>12</td>
<td>x-tricosene</td>
<td>2294</td>
<td>nd</td>
<td>44</td>
<td>hexacosane</td>
</tr>
<tr>
<td>13</td>
<td>tricosane</td>
<td>2300</td>
<td>2300</td>
<td>45</td>
<td>9,13-heptacosadiene</td>
</tr>
<tr>
<td>14</td>
<td>11- &amp; 9-methyltricosane</td>
<td>2336</td>
<td>2336</td>
<td>46</td>
<td>7,11-heptacosadiene</td>
</tr>
<tr>
<td>15</td>
<td>7,11-tetracosadiene</td>
<td>2355</td>
<td>nd</td>
<td>47</td>
<td>9,9-heptacosadiene</td>
</tr>
<tr>
<td>16</td>
<td>x,y-tetracosadiene</td>
<td>2363</td>
<td>2363</td>
<td>48</td>
<td>7-heptacosene</td>
</tr>
<tr>
<td>17</td>
<td>3-methyltricosane</td>
<td>2373</td>
<td>2373</td>
<td>49</td>
<td>5-heptacosene</td>
</tr>
<tr>
<td>18</td>
<td>7-tetracosene (quantified only in ♂)</td>
<td>2377</td>
<td>2377</td>
<td>50</td>
<td>heptacosane</td>
</tr>
<tr>
<td>19</td>
<td>6-tetracosene</td>
<td>2380</td>
<td>2380</td>
<td>51</td>
<td>8,12-octacosadiene</td>
</tr>
<tr>
<td>20</td>
<td>NI (variable in DGRP but nd in GC-MS samples)</td>
<td>nd</td>
<td>nd</td>
<td>52</td>
<td>2-methylheptacosane (or only 2-methylheptacosane)</td>
</tr>
<tr>
<td>21</td>
<td>5-tetracosene</td>
<td>2386</td>
<td>2386</td>
<td>53</td>
<td>6,10-octacosadiene</td>
</tr>
<tr>
<td>22</td>
<td>tetracosane</td>
<td>2400</td>
<td>2400</td>
<td>54</td>
<td>3-methylheptacosene</td>
</tr>
<tr>
<td>23</td>
<td>9,13-pentacosadiene</td>
<td>2451</td>
<td>nd</td>
<td>55</td>
<td>9,13-nonacosadiene</td>
</tr>
<tr>
<td>24</td>
<td>7,11-pentacosadiene</td>
<td>2460</td>
<td>nd</td>
<td>56</td>
<td>9,13-nonacosadiene</td>
</tr>
<tr>
<td>25</td>
<td>2-methyltetrasocene</td>
<td>2463</td>
<td>2463</td>
<td>57</td>
<td>2-methyloctacosane (or only 2-methyloctacosane)</td>
</tr>
<tr>
<td>26</td>
<td>6,10-pentacosadiene</td>
<td>2468</td>
<td>2468</td>
<td>58</td>
<td>5,9-nonacosadiene (or only 9-nonacosadiene)</td>
</tr>
<tr>
<td>27</td>
<td>5,9-pentacosadiene</td>
<td>2474</td>
<td>2474</td>
<td>59</td>
<td>7-nonacosene</td>
</tr>
<tr>
<td>28</td>
<td>8-pentacosene</td>
<td>2478</td>
<td>nd</td>
<td>60</td>
<td>nonacosane</td>
</tr>
<tr>
<td>29</td>
<td>7-pentacosene</td>
<td>2482</td>
<td>2483</td>
<td>61</td>
<td>8,12-triacontadiene</td>
</tr>
<tr>
<td>30</td>
<td>6-pentacosene</td>
<td>2485</td>
<td>nd</td>
<td>62</td>
<td>2-methyltriacontane</td>
</tr>
<tr>
<td>31</td>
<td>5-pentacosene</td>
<td>2492</td>
<td>2492</td>
<td>63</td>
<td>7,11-hentriacontadiene</td>
</tr>
<tr>
<td>32</td>
<td>pentacosane</td>
<td>2500</td>
<td>2500</td>
<td>64</td>
<td>hentriacontane</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IS</td>
<td>dotriacontane</td>
<td>3200</td>
<td>3200</td>
</tr>
</tbody>
</table>

"Cuticular component" column indicates the specific compound identified. "Retention index" column lists the retention indices for each compound, with male and female values separated by a comma. "NI" indicates not identified, and "nd" indicates not detected. "Bold typeface" indicates compounds not previously identified in D. melanogaster.
Table 4.2 Phenotypes and \textit{In(3R)K} genotypes for females from DGRP lines with functional \textit{Desat2} alleles. Red text indicates "mismatched" \textit{Desat2} genotype ('+' = ancestral; '-' = 16-bp deletion) and inversion status ('INV' = \textit{In(3R)K}; 'ST' = Standard karyotype). Blue background indicates "mismatched" \textit{Desat2} genotype and phenotype.

<table>
<thead>
<tr>
<th>DGRP line</th>
<th>\textit{Desat 2} genotype</th>
<th>% 7,11-C27:2</th>
<th>% 5,9-C27:2</th>
<th>Ratio</th>
<th>\textit{In(3R)K} status</th>
</tr>
</thead>
<tbody>
<tr>
<td>DGRP_31</td>
<td>+ / -</td>
<td>23.7</td>
<td>18.7</td>
<td>1.27</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_38</td>
<td>+ / -</td>
<td>19.4</td>
<td>24.9</td>
<td>0.78</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_48</td>
<td>+ / -</td>
<td>19.1</td>
<td>25.0</td>
<td>0.76</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_100</td>
<td>+ / +</td>
<td>22.3</td>
<td>32.3</td>
<td>0.69</td>
<td>INV / INV</td>
</tr>
<tr>
<td>DGRP_136</td>
<td>+ / -</td>
<td>14.7</td>
<td>33.7</td>
<td>0.44</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_309</td>
<td>+ / -</td>
<td>19.1</td>
<td>17.8</td>
<td>1.07</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_440</td>
<td>+ / -</td>
<td>20.0</td>
<td>14.0</td>
<td>1.43</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_559</td>
<td>+ / -</td>
<td>23.5</td>
<td>10.5</td>
<td>2.24</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_646</td>
<td>+ / +</td>
<td>17.3</td>
<td>24.6</td>
<td>0.70</td>
<td>INV / INV</td>
</tr>
<tr>
<td>DGRP_802</td>
<td>+ / -</td>
<td>18.5</td>
<td>23.4</td>
<td>0.79</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_732</td>
<td>+ / -</td>
<td>19.3</td>
<td>17.3</td>
<td>1.12</td>
<td>INV / ST</td>
</tr>
<tr>
<td>DGRP_367</td>
<td>+ / +</td>
<td>12.6</td>
<td>36.8</td>
<td>0.34</td>
<td>ST / ST</td>
</tr>
<tr>
<td>DGRP_776</td>
<td>+ / +</td>
<td>16.7</td>
<td>8.98</td>
<td>1.86</td>
<td>ST / ST</td>
</tr>
<tr>
<td>DGRP_509</td>
<td>+ / +</td>
<td>17.8</td>
<td>27.5</td>
<td>0.65</td>
<td>ST / ST</td>
</tr>
<tr>
<td>DGRP_235</td>
<td>+ / -</td>
<td>10.3</td>
<td>22.2</td>
<td>0.46</td>
<td>ST / ST</td>
</tr>
<tr>
<td>DGRP_551</td>
<td>+ / -</td>
<td>26.6</td>
<td>4.29</td>
<td>6.20</td>
<td>ST / ST</td>
</tr>
<tr>
<td>DGRP_105</td>
<td>+ / +</td>
<td>21.4</td>
<td>5.10</td>
<td>4.20</td>
<td>INV / INV</td>
</tr>
</tbody>
</table>
Figure 4.2 DGRP lines segregate for the female African CHC phenotype, Desat2 allele, and In3R(K) inversion status. (A) Overlaid chromatograms of African D. melanogaster CHCs (Z30 and Z53), a DGRP line with an African-like CHC phenotype (DGRP_235), and a Cosmopolitan DGRP line (DGRP_714). (B) DGRP lines with at least one ancestral Desat2 allele exhibit natural variation in the percentage of each CHC peak for the isomeric sex pheromones 7,11-C27:2 (2-Me-C26 co-elutes with 7,11-C27:2) (gray) and 5,9-C27:2 (9-C27:1 co-elutes with 5,9-C27:2) (red). (C) Box-plots of the proportion of each sex pheromone peak for DGRP and African lines according to Desat2 allele and In3R(K) genotypes. DGRP_105 and DGRP_551, which have more Cosmopolitan-like phenotypes despite having the ancestral Desat2, are indicated.
The co-segregation of the functional allele and \textit{In3R(K)} was not perfect: four of the 16 lines with the ancestral \textit{Desat2} sequence were homozygous for the standard karyotype (Fig 4.2C, Table 4.2). Furthermore, two lines with the ancestral and presumably functional \textit{Desat2} did not exhibit the African CHC phenotype, DGRP\_105 and DGRP\_551. In total, five of the 17 lines with the ancestral sequence had mismatched inversion and CHC status in females. It is possible that these lines are segregating for the inversion at low frequency and thus only the standard inversion flies were sampled for karyotyping. Alternatively, the \textit{Desat2} deletion may have occurred prior to the inversion event.

We next checked \textit{Desat2} for potentially damaging genetic variants (40,41). We identified five \textit{Desat2} alleles unique to DGRP\_105. Two variants are synonymous coding (\textit{3R\_8262545\_SNP} and \textit{3R\_8263020\_SNP}), one is a deletion causing a frameshift (\textit{3R\_8263023\_DEL}), and two are nonsynonymous coding (\textit{3R\_8263031\_MNP} and \textit{3R\_8263048\_SNP}). The frameshift and nonsynonymous variants are potentially damaging and could explain why this line produces the Cosmopolitan phenotype despite having the functional \textit{Desat2} allele. We did not find such evidence for DGRP\_551, suggesting this line may contain unknown genetic variants that inhibit the production of 5,9-C27:2.

\textit{CHC correlations and principal components analyses}

Most of the CHCs belong to homologous series in which the chain length increases by two carbons; thus, these compounds may be genetically correlated due to a shared biosynthetic pathway and the data may be confounded by multicolinearity (43). We visualized the correlations between CHCs using modulated modularity clustering (MMC) (44). The MMC algorithm clusters highly correlated variables based on the Spearman's rank correlation coefficients (\(\rho\)). As expected, some CHCs were highly correlated within each sex (Fig 4.3 and Fig 4.4; S4 Table).
Figure 4.3 MMC modules of DGRP female CHCs based on Spearman's rank correlation coefficients ($\rho$). Correlations are color-coded from +1 (dark red) to -1 (dark blue). Correlated CHCs are clustered into groups (modules). Modules (outlined in black) are arranged along the diagonal according to the average strengths of the correlations within each cluster; the most strongly correlated modules are on the top left and the weakly correlated modules are on the bottom right.
Figure 4.4 MMC modules of DGRP male CHCs based on Spearman's rank correlation coefficients ($\rho$). Correlations are color-coded +1 (dark red) to -1 (dark blue). Correlated CHCs are clustered into groups (modules). Modules (outlined in black) are arranged along the diagonal according to the average strengths of the correlations within the groups; the most strongly and weakly correlated are on the top left and bottom right, respectively.
In the first two female modules, seven dienes had strong positive correlations with each other. There was also one peak (47 – 5,9-C27:2 and 9-C27:1) that strongly negatively correlated with those dienes and the shorter-chained (≤ C25) alkanes of module 3 (Fig 4.3, S4 Table). Similarly, the module 3 shorter-chained alkanes had strong positive correlations with each other, some dienes of modules 1 and 2, and monoenes and dienes in module 5. These four modules (1, 2, 3, and 5) all had weak to moderate negative correlations with module 7, which consisted of strongly intercorrelated longer-chained (≥ C25) alkanes and dienes.

In the male CHCs, we found similar trends (Fig 4; S4 Table). Module 1 consisted of longer-chained alkanes that negatively correlated with the shorter-chained alkanes of module 2. This could also be seen between module 2 and the longer-chained monoenes in module 7. These negative correlations were a consistent trend between other long and short-chained CHCs and exemplify the tradeoff between short and long-chained compounds, as the latter are produced through the elongation of fatty acids that serve as precursors for the former.

The correlations among CHCs are plausible given the biology of CHC production. We computed PCs of genetically variable CHCs within each sex to reduce the dimensionality of the data to orthogonal PCs. The first seven and five PCs accounted for 98.00% and 98.12% of the total variation among the DGRP lines for female and male CHCs, respectively (Table 4.3, Fig 4.5, S5 Text). We hypothesized that the GWA results would provide insights into the genetic architecture underlying the MMC trends.

Table 4.3 Percent of CHC variation in the DGRP explained by PCs.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Number</th>
<th>Eigenvalue</th>
<th>Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1</td>
<td>0.0061</td>
<td>41.16</td>
<td>41.16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0043</td>
<td>29.47</td>
<td>70.63</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0021</td>
<td>14.50</td>
<td>85.13</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0009</td>
<td>6.22</td>
<td>91.35</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0005</td>
<td>3.07</td>
<td>94.42</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.0003</td>
<td>2.30</td>
<td>96.72</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.0002</td>
<td>1.29</td>
<td>98.01</td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>0.0170</td>
<td>75.52</td>
<td>75.52</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0033</td>
<td>14.57</td>
<td>90.10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.0010</td>
<td>4.59</td>
<td>94.69</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0005</td>
<td>2.04</td>
<td>96.73</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0003</td>
<td>1.39</td>
<td>98.12</td>
</tr>
</tbody>
</table>
Figure 4.5 Principal component biplots for PC1 and PC2 of DGRP CHCs. (A) Female and (C) male PC1 and PC2. (B) Female and (D) male PC1 and PC2 eigenvectors. The percent of variance explained by each PC is indicated on the x- and y-axes. In (A) and (C) DGRP lines are color-coded (S5 Text).
**GWA analyses**

We performed GWA analyses using the PCs of natural variation in individual chromatographic peaks to identify novel components of the CHC metabolic pathways in *D. melanogaster*. None of the female or male PCs were affected by the presence of the endosymbiont *Wolbachia pipientis* in some of the DGRP lines (S6 Table). Female PC1 and PC2 and male PC1 were affected by the *In(2L)t* inversion; female PC1, PC2 and PC3, and male PC5 were affected by the *In(3R)K* inversion; and the *In(3R)P* inversion only affected male PCs (PC1, PC3, PC5) (S6 Table). We corrected for these effects prior to conducting the GWA analyses. Although these inversions contain genetic variants affecting CHC production, they cannot be resolved by GWA analysis due to elevated linkage disequilibrium within the inversions (40,41) and are thus excluded from consideration.

We identified genetic variants in or near 305 (173) genes nominally (*P* ≤ 10^{-5}) associated with female (male) PCs (S7 Table). Although all of the top variants did not reach individual Bonferroni-corrected significance levels, most quantile-quantile plots indicated no systematic inflation of test statistic and a clear departure from random expectation below *P* < 10^{-5}, justifying our choice of this reporting threshold and suggesting that the top associations were enriched for true positives (S8 Fig.).

Several of the top variants associated with each PC are in or near candidate genes with plausible roles in CHC metabolism (S7 Table). In females, these include a SNP in the fourth intron of *Lipase2* (*Lip2*) associated with variation in PC1; a SNP 641 bp upstream of the cytochrome P450 gene, *Cyp49a1*, associated with PC2; SNPs in and near *Dessicate* (*Desi*), a gene previously shown to contribute to desiccation resistance in *D. melanogaster*, associated with PC4 and PC6; variants in and near two peroxidase genes, *Immune-regulated catalase* (*Irc*) and *Peroxidase* (*Pxd*), associated with PC5; and variants in three fatty acid metabolism genes (*CG14688, CG9458, CG8814*) associated with variation in PC6. Several notable candidate genes were also implicated by variants associated with PC7; *CG9801* contained the 3 most significant SNPs. One of the top associated SNPs causes a missense mutation in *Cyp6w1, Cyp4s3* and *CG5599*, a predicted dihydrolipoamide branched chain acyltransferase, were down- and up-stream, respectively, of associated variants. We did not find any associated variants within or near *Desat1* or *Desat2*. In the case of *Desat2* the ancestral allele was not tested for association, because there were so few DGRP lines homozygous for the insertion that it did not reach the MAF cutoff (0.05) for evaluation (insertion MAF = 6/169 or 0.035). Further, any effect of this variant would have been minimized by correcting for the effect of the *In(3R)K* inversion.

In males, nearly all of the variants associated with variation in PC1 were in or near *CG13091*, a putative fatty acyl-CoA reductase, of which one, a nonsynonymous coding variant (*2L_8521314_SNP*), was the most significant variant in this study (*P* = 2.19E-11) and passes the
Bonferroni-correction (S7 Table). Many of the variants in \textit{CG13091} were in perfect or near-perfect linkage disequilibrium, and thus it is not possible to discern which was/were the causal variant(s). Variants in \textit{CG13091} were also associated with variation in PC4 and PC5. Genes tagged by variants associated with PC2 include \textit{approximated (app)}, a palmitoyltransferase; \textit{PHGPx}, a peroxidase; and \textit{CG16979}, a thiolester hydrolase. Top variants in the PC3 GWA analysis included two genes predicted to be fatty acid elongases (\textit{CG18609, CG30008}) and a NADH dehydrogenase (\textit{CG8680}). The most significant variant in the PC5 GWA analysis was a nonsynonymous SNP (3R\_8220563\_SNP, $P = 1.97E-09$) in \textit{CG10097}, which also encodes a fatty acyl-CoA reductase. Interestingly, there are two cytochrome P450 genes located upstream of \textit{CG10097}, one functional (\textit{Cyp9f2}) and one pseudogenized (\textit{Cyp9f3Psi}).

\textit{Functional validation of candidate genes}
We selected 24 candidate genes with plausible contributions to various stages in the biosynthesis or turnover of CHCs and tested the effects of disruption of expression of these genes on CHC composition. All but one of these genes had publicly available transgenic UAS-RNAi lines (45). We used an oenocyte-specific \textit{GAL4} driver line, \textit{PromE(800)}-\textit{GAL4} (46), to restrict reduction in candidate gene expression to the CHC-producing cells. For the gene for which no RNAi line was available, \textit{CG10097}, we used a \textit{PiggyBac} insertion line to study the effects of this mutation on CHC composition (S9 Table) (47).

Fatty acid biosynthesis begins with the decarboxylation of malonyl-CoA and proceeds through subsequent additions of two-carbon units derived from acetyl-CoA through a series of reduction reactions mediated by enzymes of the acyl carrier protein complex. The longest chain product is the 16-carbon unit of palmitate and is released from the ACP complex by thioesterases or through a palmitoyl transferase with subsequent removal of a conjugating thiolester bond through the action of thioreductases. Fatty acids with shorter chains are generated when the hydrocarbon chain is released from the ACP complex before reaching a length of 16 carbons. Extension of the hydrocarbon chain length beyond 16 carbons requires special elongation reactions, which occur either in the mitochondria or the endoplasmic reticulum.

The candidate genes implicated by our GWA analyses that we selected for further functional assessment are annotated to encode a palmitoyl transferase (\textit{app}), fatty acyl reductases (\textit{CG13091 and CG10097}), a thiol hydrolase (\textit{CG16979}), thioredoxin peroxidases (\textit{PHGPx} and \textit{Prx6005}), fatty acid elongases (\textit{CG30008, CG18609, and CG9458}), cytochrome P450s (\textit{Cyp49a1, Cyp9f2, and Cyp4s3}), peroxidases (\textit{Irc, Pxd and Pxn}), an NADH dehydrogenase (\textit{CG8680}), and a dihydrolipoamide branched chain acyltransferase (\textit{CG5599}), which is involved in the metabolism of branched chain amino acids leucine, isoleucine and valine (42), which serve as precursors for methyl
branched alkanes (1). Disruption of expression of any of these genes by targeted RNA interference resulted in altered CHC compositions, often with sexually dimorphic effects. The alterations in CHC composition as a consequence of gene disruption were often complex and unexpected (Fig 4.6, Fig 4.7).

Fatty acyl-CoA reductases catalyze the reduction of fatty acyl-ACP to fatty alcohols which are converted to aldehydes and subsequently to hydrocarbons (48,49). *CG13091* and *CG10097* encode fatty acyl-CoA reductases, which are both expressed at high levels in males and at low levels in females (42). Inhibition of the production of these fatty acyl reductases promoted the production of longer chained CHCs. Males in which *CG13091* expression was targeted with RNAi had higher relative amounts of longer-chained CHCs in general and reduced shorter-chained monoenes and methyl-CHCs. In particular, 7-C25:1 and 7-C27:1 and 2-MeC26, and 2-MeC28 increased substantially (Fig 4.7, Fig 4.8). The increase in 7-C25:1 is of particular interest, as this compound acts as a male sex pheromone that mediates female mate choice. Females showed a similar trend but to a lesser extent; they also had some increased longer-chained dienes and, like the males, increased 7-C27:1, 7,11-C27:2 & 2-MeC26, and 2-MeC28 & 7,11-C29:2 (Fig 4.6, Fig 4.8). Both sexes of the *CG10097* mutant were similar, with higher relative amounts of longer-chained CHCs. They also had lower amounts of shorter-chained monoenes and methyl-CHCs. In the *CG10097* mutant and control females, we were able to separate the 7,11-C27:2 and 2-MeC26 peaks as well as the 2-Me-C28 and 7,11-C29:2 peaks (Fig 4.8), enabling us to infer decreased 7,11-C27:2 and 7,11-C29:2 and increased 2-Me-C28 relative to the control. Our results suggest that the fatty acyl-reductases encoded by *CG13091* and *CG10097* may be specifically associated with alkane and monoene synthesis. *Nrt*, a carboxylic acid esterase, could also catalyze the reduction of the carboxylic ester group on the fatty acyl-ACP to an alcohol.
Figure 4.6 Summary of RNAi and mutant experiments for female CHCs. UAS-RNAi target gene and the CG10097\textsuperscript{500276} mutant are indicated on the horizontal axis. CHC names and numbers are listed on the y-axis. Data are color coded to represent $P$-values ($P \leq 0.05$) from t-tests for the mean differences of the experimental and the control lines. Black = no significant change; blue = significant decrease; green = significant increase; gray = not applicable (peaks 46 and 57 split into two peaks for the CG10097 mutant).
Figure 4.7 Summary of RNAi and mutant experiments for male CHCs. UAS-RNAi target gene and the CG10097\textsuperscript{e00276} mutant are indicated on the horizontal axis. CHC names and numbers are listed on the y-axis. Data are color coded to represent P-values (P ≤ 0.05) from t-tests for the mean differences of the experimental and the control lines. Black = no significant change; blue = significant decrease; green = significant increase; gray = not applicable (peaks 46 and 57 split into two peaks for the CG10097 mutant).
Figure 4.8 Example chromatograms of oenocyte-specific RNAi knockdowns and mutants – CG13091 and CG10097. (A) and (B) PromE(800)-GAL4 x UAS-CG13091. (C) and (D) Exelixis mutant CG10097 e00276. pA = picoAmperes, IS = internal standard, ↑ CHCs significantly increased or ↓ decreased according to the individual t-tests.
The complexity of effects on CHC composition through RNAi-targeting is illustrated by the diverse effects on CHC composition of the three different fatty acid elongases encoded by CG30008, CG18609 and CG9458. In males, CG30008 may play a role in the elongation of n-alkanes and monoenes, and CG18609 may elongate precursors of longer-chained n-alkanes and 2-Me-alkanes (50). Disruption of CG30008 expression in males resulted in greatly reduced amounts of total CHCs (~2-fold). The effect was less severe in females, but the overall trend was here also towards reduced CHCs (Fig 4.9). Disruption of CG18609 decreased longer-chain CHCs in both sexes but increased 2-Me-C24 in females consistent with the preliminary studies of this gene (Fig 4.6) (50). Males also had increased total CHCs and large increases in shorter-chain CHCs (Fig 4.7). This effect was mimicked by interference with expression of Cyp9f2. This also resulted in fewer longer-chain monoenes, alkanes, and many methyl-branched CHCs. However, as for CG18609, 2-Me-C24 increased in females and other shorter-chain methyl-branched CHCs were trending upward. In males, there were overall increases, but especially in the shorter-chain CHCs. Thus, Cyp9f2 appears to be also involved in the elongation of precursors of longer-chained n-alkanes and monoenes in females. These observations suggest that CG18609 may function in the biosynthesis of CHCs in coordination with Cyp9f2. Interestingly, the CG9458 knockdown had sexually dimorphic effects on CHC production, increasing male CHCs and decreasing nearly all n-alkanes and monoenes in females. This suggests that CG9458 is critical for the biosynthesis of n-alkanes and monoenes in females.
Figure 4.9 Example chromatograms of oenocyte-specific RNAi knockdowns – CG8680 and CG30008. (A) and (B) PromE(800)-GAL4 x UAS-CG8680. (C) and (D) PromE(800)-GAL4 x UAS-CG30008. pA = picoAmperes, IS = internal standard, ↑ CHCs significantly increased or ↓ decreased according to the individual t-tests.
The three cytochrome P450s that we tested (Cyp49a1, Cyp9f2, Cyp4s3) also affected overall amounts of CHCs. RNAi knockdown of Cyp49a1 and Cyp9f2 in males led to general increases in CHCs, with a few exceptions. The increase in CHCs in males with compromised Cyp49a1 function was accompanied by a decrease in many female CHCs. Reduced expression of Cyp4s3 resulted in increased CHC abundance in both sexes. These results suggest that oxidation reactions mediated by these cytochrome P450s may regulate CHC degradation and turnover.

The complex interrelationships that give rise to variation in sexually dimorphic CHC profiles is further illustrated by RNAi interference of Irc, Pxd, and Pxn, which all have corresponding alleles associated with variation in CHC composition in the DGRP and encode peroxidases. However, interference with their expression through targeted RNAi resulted in different shifts in CHC composition. Disruption of Irc reduced the amounts of monoenes and increased the amounts of dienes in females, while increasing male CHCs, reminiscent of the effects of disruption of Cyp49a1. The effects of interference with Pxd were more complex. In females, nearly all CHCs increased, while in males many odd-chained CHCs increased, but there were decreases only in even-chained CHCs. A similar phenomenon is observed with disruption of Cyp4s3. Pxn may be important for diene synthesis because the knockdown females had decreased dienes and corresponding increased levels of longer-chained monoenes. However, in males there were decreased longer-chained monoenes, methyl-branched alkanes, and n-alkanes but increased shorter-chained CHCs. CG7724 contributes to oxidation-reduction processes and steroid synthesis (42). Disruption of CG7724 expression in females resulted in a decrease in longer-chained monoenes and methyl-branched CHCs and an increase in 2-Me-C22, 2-Me-C24, 2-Me-C25, 2-Me-C26 and 7,11-C27:2, x,y-C24:2 and x,y-C26:2. Males also showed increases in 2-methyl-CHCs but also in shorter-chain alkanes and monoenes. These results reveal a complex and dynamic network of oxidative enzymes of which the summed activity determines the sexually dimorphic composition of CHCs.

Disruption of the NADH dehydrogenase CG8680 and of CG5599, which mediates the mitochondrial breakdown of branched amino acids, resulted in remarkable increases in the total amount of CHCs in both males and females (Fig 4.6, Fig 4.7, Fig 4.9, S10 Fig., S11 Table). Individual control female and male flies produced ~1.5–2 µg and ~1–1.5 µg of CHCs, respectively; in contrast, the RNAi-CG8680 females and males produced ~5.5 µg and ~3 µg, respectively, of CHCs per fly (S10 Fig). The CG5599 knockdown caused a >4 µg increase per fly in each sex, representing ~ 3-fold (~7 µg/fly) and 4-fold (~5.5 µg/fly) increase for females and males, respectively. Thus, inhibition of central catabolic pathways resulted in an increase in CHCs, as expected. Inhibition of expression of Pxd, CG8814 and Cyp4s3 also resulted in increased levels of total CHCs in both sexes, with notable decreases in 2-Me-C26 in Pxd and Cyp4s3 males (Fig 4.6, Fig 4.7, S10 Fig.).
The gene products of *app*, *PHGPx*, and *Prx6005* may be involved in the release of fatty acids from the fatty acid synthase complex. The *app* palmitoyltransferase may be specific to methyl-branched fatty acids since we observe decreases in methyl-branched CHCs in males and females when its expression is disrupted. Male and female PCAs were here clearly separated from the controls (S10 Fig.) and in both sexes the total amounts of CHCs were slightly increased (Fig 4.6, Fig 4.7). Additionally, males had elevated longer-chained methyl-CHCs, monoenes and *n*-alkanes, while females had increased abundances of *n*-alkanes, monoenes, dienes, and methyl-CHCs of both longer- and shorter-chain lengths. RNAi targeting of *Prx6005* increased shorter-chain monoenes and alkanes in both sexes. In females, longer-chain dienes and methyl-branched CHCs were trending downward, and in males levels of 2-Me-C26 and 2-Me-C24 decreased. Disruption of *PHGPx* increased the dienes 7,11-C25:2 and 9,13-C25:2, and some longer-chained *n*-alkanes in females. Levels of many short-chained monoenes and nearly all 2-methyl compounds were elevated in males. Furthermore, disruption of *CG16979*, which encodes a gene product annotated as having thiolesterase activity, had no effect in females but caused many CHC increases in males, predominantly in monoenes. Total CHCs were also elevated (Fig 4.6, Fig 4.7). Thus, there appears to be functional specialization among thiolesterases in the biosynthesis of CHCs.

We also assessed the effects of two candidate genes, *Desi* and *Lip2*, associated with variation in CHC composition in the DGRP that affect desiccation resistance. Expression of *Desi* fluctuates in wandering *D. melanogaster* larvae in response to environmental conditions and RNAi knockdown of *Desi* in larvae leads to higher mortality (51). However, the nature of the resistance to desiccation is unknown and CHCs were not phenotyped in the larvae. In contrast to the GWA results, knocking down *Desi* had no effect on female CHCs, while the male CHCs clearly separated from the controls in the PCA (S10 Fig.). Males had increased alkanes, monoenes, and 2-Me-C28 (Fig 4.7). *Lip2* has been associated with clinal variation of life history traits in *D. melanogaster* populations in the eastern United States (52), and has triglyceride lipase activity that could regulate release of free fatty acids for incorporation in CHCs. Females in which expression of *Lip2* was targeted with RNAi had increased *n*-alkanes and increased total CHCs, whereas the males had increased 2- and 3-methyl alkanes (Fig 4.6).

Finally, we examined the effects of RNAi with predicted molecular functions of alpha-oxidation [42], *CG14688* and *CG9801*. Suppression of gene expression resulted in sexually dimorphic increases and decreases in overall CHCs. RNAi targeting of *CG14688* increased female CHCs but had more complex effects in the males, increasing shorter-chained *n*-alkanes and methyl-branched alkanes and decreasing longer-chained monoenes. Males expressing RNAi against *CG9801* had increased total CHC levels, while females of these lines had overall decreases in CHCs. In mammals, alpha-oxidation is used to break down 3-methyl-fatty acyl-CoAs. Our results suggest alpha-oxidation
may be needed for 3-methyl-branched CHC metabolism in *D. melanogaster* females or for the synthesis of longer-chained CHCs from shorter-chained precursors in males. However, we are not aware of any examples of alpha-oxidation in insects to date, and this does not explain the increases of other CHCs without 3-methylbranching in the RNAi knockdowns.

**DISCUSSION**

Using the DGRP, we report one of the most comprehensive characterizations of natural variation in insect CHCs to date. We provide unambiguous evidence for extensive heritable individual variation in the relative abundances of CHC components. Since CHCs represent the boundary between the organism and its environment and mediate social interactions while offering protection against adverse environmental effects, variation in CHC profiles may present a target for natural selection and adaptive evolution. MMC analysis shows that CHC components are not independent, but can be organized as correlated modules, which likely reflect common biosynthetic origins. Variation in the CHC profiles can be captured with a limited number of PCs, which were used as composite phenotypic values for GWA analyses. This resulted in the identification of candidate genes associated with variation in CHC composition in males and females, including several genes, which could be plausibly associated with CHC biosynthesis. Despite a lenient significance threshold, targeted gene disruption of each of these 24 candidate genes indeed affected CHC composition, often in a sexually dimorphic manner. Our results provide a new perspective and highlight the complexity of the biosynthetic and catabolic pathways that regulate the dynamics of CHC composition.

**Phenotypic characterization of CHC profiles**

We found substantial heritable natural variation in CHC composition for males and females of the inbred, sequenced DGRP lines. Several of the epicuticular compounds identified in this study have not previously been reported for *D. melanogaster* (18,37,53). These compounds separated in our GC analyses because we used a thin high-resolution column and a relatively long temperature program. Several of the newly identified monoenes and dienes had double bonds in an even carbon position, an unusual configuration in insects. While most of these new compounds represented a very small fraction of the total CHCs, two (peak 51 = 8,12-octacosadiene and peak 53 = 6,10-octacosadiene, 9-octacosene and 3-methylheptacosane) were female-specific and could potentially play a role in sexual communication.

**A segregating African CHC phenotype in a cosmopolitan population**

Surprisingly, multiple DGRP lines presented clear African CHC phenotypes and the corresponding *Desat2* genotype. The African CHC phenotype has an abundance of 5,9-C27:2 (& 9-C27:1) and lower levels of 7,11-C27:2 (& 2-Me-C26), and is present only in populations from Sub-Saharan Africa and
the Caribbean (27). Based on genotyping the 16-bp deletion/ancestral Desat2 allele, Caribbean populations are known to have spread northward into the southern United States; however, populations north of Alabama and Mississippi were thought to be nearly fixed for the Cosmopolitan deletion (54). We found that the DGRP is segregating for the African CHC phenotype. Thus to the best of our knowledge the DGRP progenitor population at the N.C. Farmers’ Market represents the northernmost population of D. melanogaster with the African Desat2 allele documented to date.

While 17 DGRP lines contain the ancestral allele that confers a functional Desat2, only 15 of these lines exhibited the African phenotype. On average, females of DGRP lines that were heterozygous or homozygous for the functional allele, regardless of inversion status, had intermediate amounts of the sex pheromone CHC peaks relative to the African lines or DGRP lines with the deletion. However, one DGRP line, DGRP_367, had more 5,9-C27:2 than either African line. Two DGRP lines (DGRP_105, ins/ ins, INV/ INV and DGRP_551, ins/ del, ST/ ST) had that functional Desat2 allele, yet exhibited the Cosmopolitan phenotype. These results are in contrast to previous reports of complete association of the ancestral allele with the African female CHC phenotype (28,55). Further analyses showed that other Desat2 polymorphisms DGRP_105 likely result in a non-functional protein. However, the other Cosmopolitan-like line, DGRP_551, is puzzling because it is heterozygous for the functional allele but homozygous for the standard karyotype. It is possible that this line is segregating for the In(3R)K inversion at low frequency and individuals with the ST/ INV karyotype were not detected, but the disconnect between the functional allele and CHC status remains unexplained since Desat2 itself does not harbor potentially damaging mutation in this line. Perhaps a polymorphism in an unknown gene in DGRP_551 interacts with the Desat2 functional allele to suppress its effects. Finally, there were four DGRP lines that exhibited the African phenotype and were homozygous for the ancestral allele, but they were ST/ ST in karyotype. One possible explanation for this “mismatching” of phenotype and genotype with the karyotype is that the Desat2 16-bp deletion occurred prior to the inversion event. This would mean that the inversion may be segregating for the Desat2 16-bp allele.

GWA analysis and functional assessment of candidate genes

We did not detect variants in the DGRP in any of the known CHC biosynthesis genes (FASN1, Desat1, eloF, DesatF, Cyp4G1) associated with CHC variation. However, we did find many novel candidate genes, and functional tests showed that disruption of expression of all tested candidate genes had significant effects on CHC production. While the mechanistic relationships between any of these genes are unknown, some share commonalities in their phenotypes when their expression is disrupted with RNAi. However, the majority of genes which encode gene products with similar molecular functions show different shifts in CHC profiles when disrupted. Thus, variation in the CHC
profile arises as an emergent phenotype from the dynamics of complex interrelated biosynthetic and catabolic pathways. The RNAi-induced shifts in CHC profiles are frequently sexually dimorphic. This could reflect different expression levels of metabolic enzymes associated with CHC production. However, we cannot exclude the possibility that differential effectiveness of RNAi in males and females may contribute to apparent sexual dimorphism.

We note that the 24 candidate genes on which we focused represent only a subset of all candidate genes associated with variation in CHC composition. Many of these genes may directly or indirectly affect CHC composition through as yet unknown mechanisms. Furthermore, it is important to note that the RNAi and mutant experiments test for effects at the level of genes and are only proxies for the effects of segregating natural variants in these genes. Nevertheless, our results provide a framework for future studies of the mechanisms that regulate CHC composition and their adaptive potential regarding cold/heat tolerance and desiccation resistance, and pleiotropic effects on chemical communication and mate choice.

MATERIALS AND METHODS

*Drosophila stocks and phenotyping*

DGRP and African (Z30 and Z53) lines were reared in vials containing cornmeal-molasses-agar medium at 25°C, 75% relative humidity, a 12:12-h light-dark cycle, and a controlled adult density of 10 males and 10 females. The parental generation was allowed to lay eggs for three days. Upon eclosion, virgin males and females were separated and placed into new vials containing the same medium and aged for four days.

The flies from each line were separated into at least two samples of five flies each per sex. On average, three samples were collected for each. To avoid cross-contamination of cuticular lipids a fresh tissue was placed on the carbon dioxide pad and the flies were handled with acetone-washed titanium forceps at each round of sorting. Flies were placed in 2 mL glass auto-injection vials with a Teflon cap and were flash frozen. All samples were stored at -30°C until CHC extraction. We collected samples from 169 and 157 DGRP lines for females and males, respectively (1078 total samples). All lines were reared simultaneously and DGRP lines that did not produce sufficient offspring for CHC analysis were excluded to avoid any block effects of rearing. For the two African lines, Z30 and Z53, we reared and collected 5 samples for each sex.

*Quantification and identification of cuticular hydrocarbons*

Cuticular lipids were extracted from each sample using 200 µl of hexane containing an internal standard (IS, 1 µg n-C32) with gentle swirling for five minutes. The flies were briefly extracted a second time with 100 µl of hexane (free of internal standard). After each wash the extract was
transferred to a 300 µl conical glass insert. The extract was dried using a gentle stream of high-purity N₂ and re-suspended in 50 µl of hexane. The samples were immediately processed using gas chromatography or stored at 4°C (no longer than one day) until processing.

The cuticular lipid extracts were analyzed using an Agilent 7890A gas chromatograph with a DB-5 Agilent capillary column (20 m x 0.18 mm x 180 µm) and a flame ionization detector (FID) for quantification. We introduced 1 µl of sample using an Agilent 7683B auto injector into a 290°C inlet operated in splitless mode. The split valve was turned on after 1 min. The oven temperature program was as follows: 50°C for one min, increased at 20°C/min to 150°C, and increased at 5°C/min to 300°C followed by a 10 min hold. Hydrogen was used as the carrier gas at constant flow (average linear velocity = 35 cm/sec) and the FID was set at 300°C.

Selected samples were analyzed for chemical identification in a 6890N GC system (Agilent) coupled with a 5975 mass selective detector (MSD) (Agilent) and equipped with a DB-5 (20 m x 0.18 mm x 0.18 µm) column (Agilent). Helium was used as carrier gas at 33 cm/s average linear velocity. Injection and temperature settings were identical to the settings described above, and the transfer line was maintained at 300°C. Positive electron ionization at 70 eV with default temperature settings (ion source at 150°C, quadrupole at 230°C) were used for the MSD. Ions were detected in scan mode in the range of 33–650 u at 1.23 scan/s scan rate. Compounds were identified based on their mass spectra in comparison to those in the reference library (Wiley 7th/NIST 05) and based on comparison of their Kovats indices and fragmentation patterns to already published Drosophila CHCs (53,56). The position of the double bond was not confirmed and chirality was not determined for any of the CHCs.

All chromatograms were analyzed using Agilent ChemStation software. For quantification individual peak areas were obtained for 42 and 60 male and female CHC containing peaks, respectively (some of the peaks contained multiple CHCs). Response factors were not determined for individual components. To account for natural variation in body size and absolute amounts of CHCs between the lines, the data were represented as proportions by dividing each peak area by the sum of all integrated peaks.

**Statistical and quantitative genetic analyses**

We partitioned variation of each CHC peak into genetic and environmental components using an ANOVA model of form \( Y = \mu + L + \varepsilon \), where \( Y \) is phenotype, \( \mu \) is the overall mean, \( L \) is the random effect of line, and \( \varepsilon \) is the residual. We estimated variance components using restricted maximum likelihood and computed the broad-sense heritability \( (H^2) \) of each CHC peak as \( H^2 = \sigma^2_L/(\sigma^2_L + \sigma^2_{\varepsilon}) \), where \( \sigma^2_L \) is the among-line variance component and \( \sigma^2_{\varepsilon} \) is the error variance. All analyses were performed with version 9.3 of the SAS System for Windows (2013 SAS Institute Inc.).
A majority of CHCs belong to homologous series in which the chain length increases by two carbons; thus these compounds may be genetically correlated due to shared biosynthetic pathways and the data may be confounded with multicollinearity (43). We visualized the correlations between CHCs using modulated modularity clustering (MMC) (44). The MMC algorithm clusters highly correlated variables based on the Spearman’s rank correlation coefficients ($\rho$). In order to take these correlations into account, we conducted principal components (PC) analysis on the variance-covariance matrices for the male and female CHC line means. For each analysis we included only CHC peaks that had an estimated $H^2 \geq 0.25$. We retained PCs explaining greater than 1% of the variation for subsequent GWA analysis. PCA was conducted in JMP Pro10 (2013 SAS Institute Inc.).

**Genome-wide association (GWA) analyses**

We conducted a GWA analysis for each CHC PC, separately for males and females. The DGRP lines are segregating for *Wolbachia* infection status and for the following common inversions: *In(2L)t*, *In(2R)NS*, *In(3R)P*, *In(3R)K*, and *In(3R)Mo*. We performed GWA studies in two stages. In the first stage, we adjusted the line means for the effects of *Wolbachia* infection and major inversions. We then used the adjusted line means to fit a linear mixed model in the form of $Y = Xb + Zu + \epsilon$, where $Y$ is the adjusted phenotypic value, $X$ is the design matrix for the fixed SNP effect $b$, $Z$ is the incidence matrix for the random polygenic effect $u$, and $\epsilon$ is the residual. The vector of polygenic effects $u$ has a covariance matrix in the form of $A\sigma^2$, where $\sigma^2$ is the polygenic variance component. We fitted this linear mixed model using the FastLMM program (version 1.09) (57). We performed these single marker analyses for the 1,883,938 (females) and 1,912,894 (males) biallelic variants (SNPs and indels) with minor allele frequencies $\geq 0.05$ whose Phred scale quality scores were at least 500 and genotypes whose sequencing depths were at least one and genotype quality scores at least 20 (41). All segregating sites within lines were treated as missing data.

**RNAi and mutant candidate gene validation experiments**

We selected candidate genes with available mutations and RNAi knockdown constructs to test for effects on CHC production based on FlyBase annotations. We obtained lines with RNAi knockdown constructs from the Vienna *Drosophila* RNAi Center (VDRC) and crossed them to the oenocyte-specific GAL4 driver, *PromE(800)-GAL4* (58,59). We tested knockdown constructs and their co-isogenic controls (F1 individuals from crosses of the empty vector strain to *PromE(800)-GAL4*) for 23 genes (S9 Table). Since no RNAi knockdown line was available from VDRC for *CG10097* identified in the male GWA analysis, we obtained and tested for this gene a *PiggyBac* insertion mutant from the Harvard Exelixis Collection (47) along with the *w^{118}* control line. From each cross and mutant line, we collected and aged both male and female virgins and analyzed the CHCs in the same manner as described for the DGRP flies. The analysis of CHCs using GC was also the same. However, for these
lines, instead of calculating the proportion that each peak contributed to the total chromatogram, we used the internal standard to calculate the amount of each CHC present in the sample (ng/fly) with the assumption that body sizes between the control and RNAi knockdown or mutant were not significantly different. These measures provide a more quantitative measure of CHCs and capture differences that proportion data may not resolve. PC analyses and t-tests pairing the test lines with the controls were conducted on these data. We also calculated the mean total amount of CHCs (ng/fly); we used the more conservative Cochran and Cox test which assumes unequal group variances to assess statistically significant effects on CHC composition. We also present these data as µg/fly in S10 Fig. t-tests were conducted in SAS v. 9.4 (2013 SAS Institute Inc.).

Acknowledgements
We thank Rick Santangelo for technical support. We also thank Drs. Akihiko Yamamoto and Richard Lyman for maintenance of fly lines and Drs. Joel Levine and Pat Estes for sharing the PromE(800)-GAL4 line. We acknowledge the Vienna Drosophila RNAi Center and the Exelixis Collection at the Harvard Medical School.

Funding statement
This work was funded by NIH R01 GM45146 and R01 GM59469 to T.F.C.M. and R.R.H.A. and the Blanton J. Whitmire endowment at NCSU (to C.S.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions
LMD reared and collected all DGRP lines and transgenic lines, extracted all CHCs, and analyzed all CHCs using GC and integrated all chromatograms for data collection. KB analyzed CHCs using GC-MS. WH and LMD conducted statistical analyses. CS developed and tested the GC methodology. LMD, TFCM, and CS conceived and designed experiments. LMD, CS, KB, WH, RRHA, and TFCM wrote the manuscript.

REFERENCES


CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS

CONCLUSIONS

For this dissertation I examined natural variation in two insect cuticular traits, body pigmentation and cuticular hydrocarbon composition. These traits showed substantial amounts of heritable variation among the DGRP lines. This was consistent with other studies conducted on the DGRP [1–6]. From the GWAS conducted on DGRP trait variation, similar trends have arisen from my research.

First, the genetic architecture of variation for each trait was more complex in the DGRP population than in populations used in previous gene mapping studies. This was likely due to the increased number of lines measured and genetic variants segregating in the DGRP population. QTL studies were conducted by creating, phenotyping, and genotyping F2 crosses from two lines with divergent phenotypes. These studies were limited by the amount of genetic variation segregating within the two lines and thus may have missed other genetic variants within the whole population from which the lines were derived [7].

Second, I identified genetic variants in and near both known and novel candidate genes associated with natural phenotypic variation [8]. In Chapter 2, I showed that genetic variants in the regulatory regions of bab1, e, and t were associated with variation in proportions of dark melanin covering tergite 5 and tergite 6. In Chapter 3, I found that variants in Mgl were associated with variation in the intensity of the melanin on tergite 6. In Chapter 4, I identified elongase genes and one of these had previously been shown to affect CHC composition in preliminary studies. The associations of genetic variants in these genes provide a proof-of-principle that the DGRP population can be used to map genetic variants associated with natural phenotypic variation.

Many of the novel candidate genes containing or near associated genetic variants are often in modular but interacting genetic networks [9]. This genetic heterogeneity is predicted by quantitative genetic theory [10]. For example, in Chapter 2, I showed that variants associated with pigmentation traits are located in different genes in the dpp, Egfr, and wg signaling pathways. Pigmentation was also affect by chitin synthesis and cuticle protein genes [8]. Genetic variants within each pathway may independently affect a phenotype or they may have epistatic interactions that lead to unexpected phenotypes. Of the 24 genes tested in CHC validation experiments, most were never previously tested for effects on CHC composition and many have few known functions. Given that previous QTL mapping studies of CHCs were not able to resolve the effects of particular genes, these finding contribute to our basic knowledge of CHC biosynthesis and lay the foundation for future experiments.

Third, associated genetic variants are often in or near genes with pleiotropic effects. In Chapter 3, three genes that were shown to affect the intensity of body coloration were previously
reported to be involved in general vesicle dynamics that also operate in renal and neuronal function. Additionally, the aforementioned signaling pathways, dpp, Egfr and wg, have been well-studied for their roles in D. melanogaster wing morphogenesis. These results highlight the interconnectivity of the genome and the deployment of the modular "genetic toolkits" in different developmental circumstances. The ability to do so likely lies in the genes’ regulatory regions.

A majority of the genetic variants identified in my studies were located in intronic or intergenic regions. Some of these regions were known gene regulatory elements. These results implicate cis-regulatory evolution, which likely limits negative pleiotropic effects, as a major contributor to phenotypic variation within the DGRP population. This is consistent with much of the recent work in the field of evolutionary developmental biology (evo-devo) [11,12].

**FUTURE DIRECTIONS**

For each of my studies, I used RNAi crosses or mutant analyses to test the effects of reduced candidate gene expression on each phenotype. However, these gene-level tests served as proxies for testing allele-specific effects. In order to do this, future studies could conduct allele-specific replacement within DGRP lines to assess the causative effects of individual SNPs on phenotypic variation. The recently developed CRISPER-Cas genome editing system is making this type of allelic substitution more feasible than ever before [13].

A common missing link for understanding the biosynthesis and deposition of pigmentation precursors and cuticular hydrocarbons is the mechanism of transport from the epidermal cells to the cuticle and oenocytes to the epicuticle, respectively. Some pieces of these puzzles have been elucidated. For example, lipophorin is known to transport CHCs from the oenocytes internally through the hemolymph, but the mechanisms by which the CHCs exit the cell and are distributed to different regions on the epicuticle are not clear [14]. My study of natural variation in cuticular hydrocarbons yielded many candidate genes that should be further investigated for roles in CHC transport and deposition.

Both of my studies on pigmentation identified genes that may function in the deposition of melanin precursors or the uptake of melanin producing enzymes from the cuticle by the epidermal cells. These include pinstripe (CG7852), Trim9, kirre, and spz5. These genes could be investigated further using in situ hybridization to assess if they are expressed in the epidermal cells during pigmentation development in the late pupal and early adult stages.

Another aspect of the pigmentation studies that remains unclear is the transport of tyrosine-derived glucosides through the hemolymph and into the epidermal cells [15,16]. These catecholamine-conjugates are not mentioned in most contemporary studies of D. melanogaster pigmentation. However, studies on other insects from older literature showed that conjugates are accumulated in the hemolymph during the mid- to late-pupal stages in preparation for cuticle
development and sclerotinization/melanization [15–20]. Future studies should assess if D. melanogaster uses these catecholamine-conjugates to store the pigmentation precursors.

CHC composition and body pigmentation are often subject to natural and sexual selection. The genes identified by my studies could be studied for their contributions to phenotypic evolution in other Drosophila species. My results provide a basis upon which to generate hypotheses about adaptations such as cold/heat tolerance and desiccation resistance, mate choice and speciation, and other phenotypes such as immune response. There are now 23 Drosophila genomes available through the National Center for Biotechnology Information (ncbi.nlm.nih.gov); these valuable resources offer the possibility for comparative genetics and bioinformatics studies.

Finally, a future study of the DGRP should investigate desiccation resistance. CHCs are well known to contribute to desiccation resistance [21,22]. However, variation in cuticle pigmentation may also contribute to variation in desiccation resistance [23]. While many studies of Drosophila report darker pigmentation in higher altitudes and latitudes, some populations show the opposite trend [24–26]. For example, D. santomea is an island endemic species that has extremely light pigmentation with no dark melanin and lives in montane cloud forests [27]. While D. santomea's lowland-dwelling, sister-species D. yakuba sports a dark black posterior abdomen [28]. Studies of Drosophila from India suggest that these inverted melanization clines may be due to adaptation for water-regulation [29–31]. In these studies, darker lowland species that experience lower humidity had lower rates of water loss. Similar trends have also been seen in Brazilian Drosophila [32]. Characterizing desiccation resistance and conducting GWA analyses in the DGRP may provide genetic cross-links between CHC composition and pigmentation traits.

REFERENCES


