CHANGES THROUGHOUT A HOX REGULATED NETWORK DRIVE BODY PLAN EVOLUTION IN *DROSOPHILA*

By

Yang Liu

BS, Nankai University, 2011

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy in Biological Sciences

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Yang Liu

It was defended on

July 19th, 2018

and approved by

Dr. Karen M. Arndt, Professor, Department of Biological Sciences

Dr. Gerard L. Campbell, Associate Professor, Department of Biological Sciences

Dr. Jon P. Boyle, Associate Professor, Department of Biological Sciences

Dr. Peter Andolfatto, Associate Professor, Department of Ecology Evolution and Behavior, Lewis-Sigler Institute for Integrative Genomics, Princeton University

Thesis Advisor: Dr. Mark Rebeiz, Associate Professor, Department of Biological Sciences

CHANGES THROUGHOUT A HOX REGULATED NETWORK DRIVE BODY PLAN EVOLUTION IN *DROSOPHILA*

Yang Liu, PhD

University of Pittsburgh, 2018

A recurring discussion in evolution biology is how the anterior-posterior animal body axis organization was diversified throughout the animal kingdom. This question has been a major focus of the field of evolutionary-developmental biology. As this discipline grew, a general hypothesis was developed which speculated that *cis*-regulatory element (CRE) evolution is a major driver of the evolution of form by altering the expression patterns of developmentally important genes. Numerous classical studies revealed that differences in body segmental identity correlated with expression shifts in the developmentally important Hox transcription factor genes. However, the inaccessibility of genetic crosses between distantly related taxa, and the complexity of Hox regulatory mechanisms built a barrier to directly implicate and pinpoint the evolutionarily relevant *cis* changes underlying such body plan differences. One further complication is that Hox genes represent just one factor in vast regulatory networks, throughout which causative variation may have accumulated. To address these problems, I investigated an evolved Hox-regulated trait in *Drosophila* — differences in abdominal pigmentation that exist between two closely related crossable species, *Drosophila* (*D.*) *yakuba* and *D. santomea*. By applying analyses of introgression lines, gene expression, transgenic reporters, and CRISPR/Cas9-based complementation tests, I have elucidated evolutionary changes throughout a Hox-regulated network in *D. santomea*: The top level Hox transcription factor *Abd-B* has evolved a temporally

and spatially restricted change in expression, which was accompanied by the gain of expression of another factor *pdm3,* which suppresses pigmentation. In three network terminal pigment-producing enzyme genes, the loss of *yellow* was attributed primarily to upstream changes, while the gain of *ebony* and the loss of *tan* resulted from changes in their CREs. I have identified most of these genes' evolutionary relevant CREs, confirmed their contribution to the phenotype, and investigated their epistatic interactions. I propose that the picture I have derived illuminates the genetic basis of body plan evolution on macroevolutionary scales, in which Hox genes evolve in unison with other loci that span vast gene regulatory networks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

PREFACE

About ten years ago, when I just started my undergraduate career, I had the chance to see the movie Jurassic Park. It moved me. Not only through the amazing action scenes, but also by the ideas it raised about evolution, genome engineering, and the respect we must have for nature. It was my first real exposure to the vivid concepts of evolution and genetics. Later, my undergraduate experience in both plant genetics and zebrafish bioengineering offered me opportunities to see how genetic experiments could be applied in model organisms, which inspired me to apply to graduate school. I was really surprised how research in evolution and development could link together evolutionary changes with the dramatic phenotypes that I found so inspiring. Upon joining the Rebeiz Lab, I came to understand the mechanisms of morphological evolution and the complexity of Hox genes, which continued to surprise me with how complex and rapid the evolution of traits can be. Now, after six years of research, I am really proud that, under the supervision of, and with the helpful advice and encouragement from Mark, I am finally able to present my thesis concerning how a Hox regulated network evolved to drive a drastic body plan difference between two closely related *Drosophila* species.

ACKNOWLEDGEMENT

I must thank my mentor and supervisor Dr. Mark Rebeiz. This work would not have been finished so comprehensively without him. The thing that first impressed me about Mark is that his ideas are always creative, and our discussions about my project would always generate rational new ideas and great methods, which continuously and progressively improved my project. Dr. Rebeiz

entrusted me with a great degree of scientific freedom to explore this problem. I still remember years ago when he asked me whether we should integrate our observations on Hox gene evolution with the other pigmentation gene changes I had found, or if we should divide these into different projects to have more papers. I said that I wanted to put them together. As we are now submitting this finalized work for publication, it turns out that this integration has yielded a much more complete picture of how genetic interactions evolve within a gene regulatory network, which is an important theme of my thesis. Even in the best case scenario during the course of a PhD, there are countless times that one is stuck by the difficulties and disappointment associated with unsuccessful experiments. Every time I was down or came to crossroads, Mark encouraged me to stay positive. Without him, I would have given up a thousand times.

I also must thank the many members of the lab, both past and present. I especially want to acknowledge our previous technician Ali Ordway, who taught me most of the dissection strategies and *in situ* hybridization processes that were critical to this project. I want to thank Dr. William Glassford, a former graduate student in the lab, for inspiring discussions of our respective projects. I must also recognize current lab members, Sarah Smith and Donya Shodja, for their considerable personal support during my time in the lab, as well as sharing the excitement of our research results. I want to thank Eden McQueen, Dr. Gavin Rice, Dr. Ben Vincent, and Aaron Novick for their thoughtful input and opinions on my research and its presentation in multiple forms. I am particularly thankful for our new lab members Iván Méndez González and Sarah Petrosky who will continue research on *Drosophila* abdominal pigmentation. I'm confident that more exciting results will be illustrated! Finally, I must recognize our current and former lab managers, Brandon Small and Stephanie Day for their upbeat personalities and the ability to make our lab feel like a home.

This work would not have been possible without the scientific support of collaborators and the scientific community of the department. I am grateful to our collaborator Dr. Stern for his great

support in generating Introgression and CRISPR fly lines, which has been fundamental to my entire project. In addition, his critical insight and discussions have greatly improved the project. I must also acknowledge the close discussions with collaborators from the Andolfatto Lab, especially Clair Han and Dr. Andolfatto. The sharing of results, especially the population analyses offered invaluable input to my thesis. I want to thank Cathy Barr, our graduate administrator in the department, who repeatedly offered help with registration, applications, and solving almost all the problems that I've encountered during my graduate life. I am indebted to my committee, Dr. Karen Arndt, Dr. Gerard Campbell, Dr. Jon Boyle and Dr. Peter Andolfatto, who all provided thoughtful advice on my project and carefully contributed to my career development.

I'd like to thank all my friends, including people from my previous undergrad university, my high school, from Pittsburgh, and online. Thanks for sharing happy moments, good movies, and touching music with me to keep my life colorful.

Last, I must thank my parents. They are both teachers, but they never simply taught me the answers. Instead, they taught me to have good study habits, encouraged me to think creatively, and continued to challenge me to work as hard I could. They encouraged me to have hobbies but be sure to focus on my work first. Their great support has been invaluable to my success in graduate school.

Yang Liu

July 10, 2018

1.0 INTRODUCTION TO THE EVOLUTIONARY STUDY OF DEVELOPNMENT, GENE REGULATION AND ANIMAL BODY PLANS

SUMMARY

A growing field of study in evolutionary biology is concerned with how morphologies evolve at the molecular and genetic levels. Capturing the molecular mechanisms of morphological evolution requires an understanding of how morphologies are built, including knowledge of development (Raff, 1996). In the evolutionary-developmental biology field (evo-devo), thanks to decades of systematic case-studies and the improvement of genetic manipulative approaches, much work has strengthened the theory that alterations in non-coding *cis*-regulatory regions of genes represent a major driver of the evolution of morphologies (Carroll, 2008; Carroll et al., 2005; Davidson, 2006; King and Wilson, 1975; Peter and Davidson, 2015; Stern, 2000). In particular, the body patterning Hox genes were a major discovery that revealed a surprising degree of conservation in how body plans are established across the animal kingdom (Khila et al., 2009; Lewis, 1978; Stern, 1998). This discovery was accompanied by many case studies that implicated changes in Hox gene expression in the divergence of animal body plans (Averof and Patel, 1997; Cohn and Tickle, 1999; Damen et al., 1998). However, it has been difficult to measure the relative contribution of Hox genes to morphological evolution and directly pinpoint the regulatory changes responsible for their expression differences. *Drosophila* abdominal pigmentation patterns vary greatly between species and among populations. Its pathway has been extensively studied at the biochemical (Wright, 1996), and genetic levels (Kopp et al., 2000; True et al., 2005; Wittkopp et al., 2002a). Furthermore, pigmentation of *Drosophila melanogaster* and its close relatives is a

Hox-regulated trait (Jeong et al., 2006). Thus, *Drosophila* abdominal pigmentation has become an ideal system to trace the connection between genotype and phenotype, and to study body plan evolution. In this chapter, I will introduce how *cis*-regulatory mutations facilitate rapid evolution of animal form, the correlations between the alterations in Hox genes and morphological evolution of macroevolutionary differences (Raff, 2000; Reznick and Ricklefs, 2009) that reside above the species level. I will introduce *Drosophila* abdominal pigmentation as a promising model to study differences between closely related species and its potential to link microevolution to macroevolution.

1.1 EVO-DEVO AND THE *CIS*-REGULATORY HYPOTHESIS

Early in the 1960s, scientists had already discovered the amino acid sequence conservation of blood proteins between human and chimpanzee (Washburn, 1963). This was further confirmed by electrophoretic and sequencing comparisons of a series of proteins between the two species (King and Wilson, 1975). However, despite the numerous traits that differ between human and chimpanzee anatomies (Bourne, 1969), the seminal study by King and Wilson found that the proteins of these two species were quite similar, stimulating the authors to suggest that the molecular basis of their morphological evolution is to a large extent dependent on the non-coding regulatory regions of genes (King and Wilson, 1975).

Another clue to the molecular basis of morphological evolution came from the body patterning Hox genes, which play critical roles in specifying identity of repeated segments along the body axis. Precise expression patterns of Hox genes determine body segment identities such as body segment color, the presence of appendages, and their segment-specific morphologies. Laboratory induced mutations in Hox genes result in homeotic transformation of one body segment's identity into that of another, such as the transformation of antennae into legs or

hindwings into forewings (Deutsch and Mouchel-Vielh, 2003; Gellon and McGinnis, 1998; McGinnis and Krumlauf, 1992). Several decades after the Hox gene complexes were discovered (Bridges and Morgan, 1923), their coding sequences were recognized to be conserved in a wide diversity of animal taxa (McGinnis et al., 1984) and in some cases the coding regions could be swapped between species, suggesting that their protein-coding functions remain stable over long evolutionary distances (Malicki et al., 1990). Additionally, the expression patterns of Hox genes exhibit a co-linear relationship with their relative arrangement on the chromosome (McGinnis and Krumlauf, 1992). These observations stimulated studies of the roles played by Hox genes during evolution, which revealed shifts in their expression, which correlate with morphological differences in body segment specialization (Averof and Patel, 1997; Burke et al., 1995). These observations implicated the evolution of their gene regulatory mechanisms.

Later, when a growing amount of molecular, developmental and genetic data started to boom in the following decades, the “bridge” to link the molecular mechanisms and morphological evolution was gradually being built. From this emerged theoretical momentum that changes in non-coding *cis*-regulatory elements **(CREs)** of developmentally relevant genes leads to expression differences which result in divergent forms (Carroll, 2008; Carroll et al., 2005; Davidson, 2006; Peter and Davidson, 2011; Peter and Davidson, 2015; Stern, 2000). In that bridge, “gene expression alteration” is the key connection between CRE changes and morphological evolution (**Figure 1.1A**). In the following subsections, I will illustrate how early examples in the field correlated gene expression shifts with morphological evolution (subsection 1.1.1), and then describe how CREs are structured, and CRE mutations can alter gene expression (subsection 1.1.2). This will be followed by examples in which such CRE mutations have been demonstrated in populations and between species (subsection 1.1.3).

1.1.1 Gene expression changes correlate with differences in morphology between species

In the animal kingdom, it is obvious to see morphological evolution, such as the number of body segments in snakes (Cohn and Tickle, 1999), the morphologies of insect legs (Khila et al., 2009), stickleback fish pelvic reductions (Shapiro et al., 2004) as well as the staggering diversity of butterfly wing patterns (Brakefield et al., 1996; Carroll et al., 1994; Reed et al., 2011). Numerous studies in the evo-devo literature have found relationships between these diverse morphologies and corresponding gene expression differences.

A representative case is the investigation of the *Manx* gene, which encodes a zinc finger protein involved in the formation of tails in Ascidian species, such as the sea squirt (Swalla and Jeffery, 1996). A comparative analysis of *Manx* expression showed that tailed Ascidians express this zinc finger transcription factor in the tail primordium, while this gene is down regulated in tailless Ascidians. In hybrid embryos of tailed and tailless species, *Manx* expression was restored, which correlated with the presence of a tail in these animals. Additionally, knocking down *Manx* expression led to the inhibition of tail development. All of these lines of evidence suggested a role for the evolution of *Manx* expression in the formation of this feature.

Work in rapidly evolving butterfly wing patterns has also provided striking correlations between gene expression and phenotype that lent support to the cis-regulatory hypothesis. First, the expression of the transcription factor *Distal-less* (*Dll*) was reported to mark the developmental organizer of butterfly wing eyespots (Carroll et al., 1994). Subsequent studies on the developing eyespots of butterflies suggested that additional genes, *engrailed*/*invected* and *spalt* correlate with the rings of color surrounding eyespots. The alteration of eyespot color matched the shift of the above transcription factors' expressions (Brunetti et al., 2001). Later analyses of red pigments

in the distantly related *Heliconius* genus revealed another transcription factor gene named *optix* whose expression correlated with striking differences in red pigment patterns among populations (**Figure 1.1 B-D**) (Reed et al., 2011).

Correlations between gene expression and evolutionary differences are not limited to morphological differences. One example is the comparison of *FoxP2* among song-learning and non-learning birds. *FOXP2* is an important gene for the development of language, and human families that have a mutation in this gene have a severe language disorder (Lai et al., 2001). The song-learning birds such as black-capped chickadee and zebra finch show higher levels of FOXP2 in the vocalization center of the brain than was found in non-learning birds, i.e. ringdoves (Carroll, 2005; Haesler et al., 2004). Interestingly, the coding sequences between the two kinds of birds don't share any fixed amino acid substitutions, and the striking degree of conservation among zebra finch, mouse and human FOXP2 protein further supported the notion that evolutionary changes in the regulatory region of the gene may have played a role in these traits (Haesler et al., 2004).

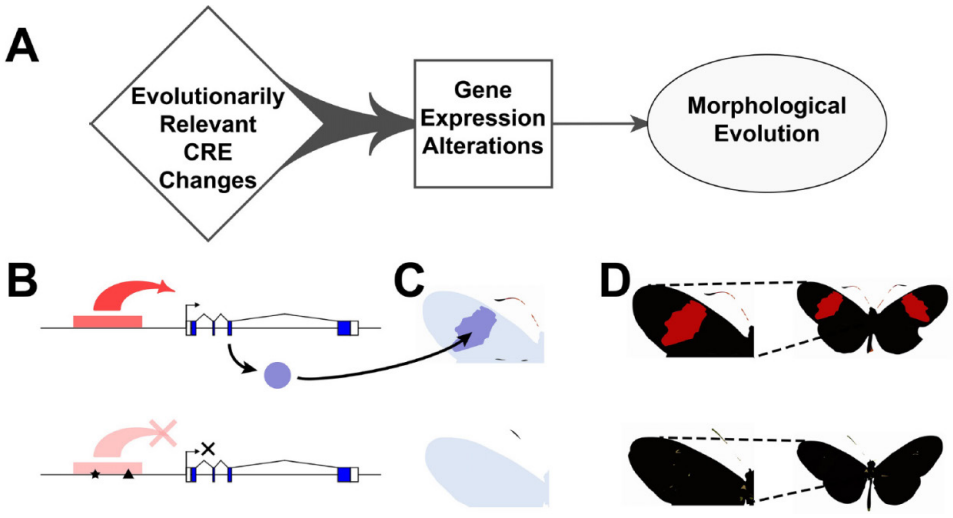

Figure 1.1 The *cis*-regulatory hypothesis of morphological evolution. (A). The bridge between evolutionarily relevant CRE changes and morphological evolution through altering gene expression patterns. (B, C). CRE evolution and gene expression alteration. The red block represents the CRE responsible for driving expression of a gene important for red pigment formation in pigmented areas of the butterfly wing. The expression pattern of the gene is shaded dark purple in (C). Black star and triangle in (B) mark evolutionarily relevant mutations that disrupt the wing CRE activity. (D) shows the morphological difference between the two butterfly wing patterns, driven by the expression changes in (C).

In addition to non-coding regulatory variants, coding sequence changes also contribute to the evolution of phenotypes. Examples of traits caused by coding sequence changes include animal fur color (Mundy, 2005), the albinism of cavefish (Protas et al., 2006), and leaf shape evolution in *Arabidopsis* (Vlad et al., 2014; Vuolo et al., 2016). One example that thoroughly illustrated coding sequence evolution is the investigation of herbivore insects who can ingest toxic plant cardenolide compounds, sometimes sequestering these toxins for their own defense (Zhen et al., 2012). Cardenolides are generated by some toxic plants and bind the Na^+, K^+-ATPase α1 channel subunit that is required to set up membrane potentials (Dobler et al., 2011). However,

some herbivore insects have evolved resistance to cardenolides and can even utilize the compounds to protect themselves from predators (Duffey and Goetz, 1980; Nishida, 2002). Zhen et al. performed a comparative analysis of Na^+, K^+-ATPase α1 subunit across three orders of insects, which demonstrated that cardenolide resistance-conferring amino acid substitutions occurred in multiple independent linages but were highly clustered. Also, in some resistant species, ATPα1 duplications were observed, suggesting that one copy could maintain its ancestral function in case the resistant amino acid substitutions are deleterious to the insect. This was implicated to be a way to avoid deleterious effects and make the evolution less constrained.

When one examines molecular alterations during evolution, it is important to consider the degree of pleiotropy, which refers to the number of phenotypic differences or tissues affected by a genetic change. Because coding sequence changes likely affect the protein product in every tissue where it appears, such changes would be highly pleiotropic. In the above examples, the encoded proteins often have a unique physiological function, and duplications also provide a way to decrease the pleiotropic effects of modifying these coding regions. However, genes related to body pattern formation and development are always highly pleiotropic. Developmental patterning genes are often initially deployed early during embryonic development (Stearns, 2010) and are thus likely integrated into many gene regulatory networks that span the life of the animal (Arnone and Davidson, 1997). As a result, coding changes in this class of gene are predicted to have deleterious effects, rendering this path less frequently traversed during evolution. Next, I will introduce how genetic alterations affect gene expression patterns and place these changes in the context of gene regulatory networks.

1.1.2 The function and evolution of *cis* regulatory elements (CREs)

To regulate gene expression, CREs can act in two ways—activating or repressing transcription. Activating CREs, which are sometimes called enhancers, are discrete elements of non-coding *cis*-acting DNA sequences that can be bound by specific transcription factor proteins to activate transcription by interacting with the promoter of a gene (Banerji et al., 1983; Blackwood and Kadonaga, 1998; Pennacchio et al., 2013). Likewise, a silencer is a CRE that represses transcription. Genes often contain multiple discrete CREs (**Figure 1.2A**), which are responsible for their expression in multiple diverse tissues. When and where these CREs are bound and activated or repressed result in differential expression in time and space (**Figure 1.2B-E**).

Of note, CREs are scattered across the majority of the genome, especially in humans (Waterston et al., 2002). However, unlike protein coding sequences, the CRE sequence code is still poorly understood, making it difficult to decipher the non-coding regulatory regions and confirm functional CREs based on computational analysis alone (Pennacchio et al., 2013). Nonetheless, some predictions of CRE function or presence are possible by using sequencing comparisons that measure conservation or divergence.

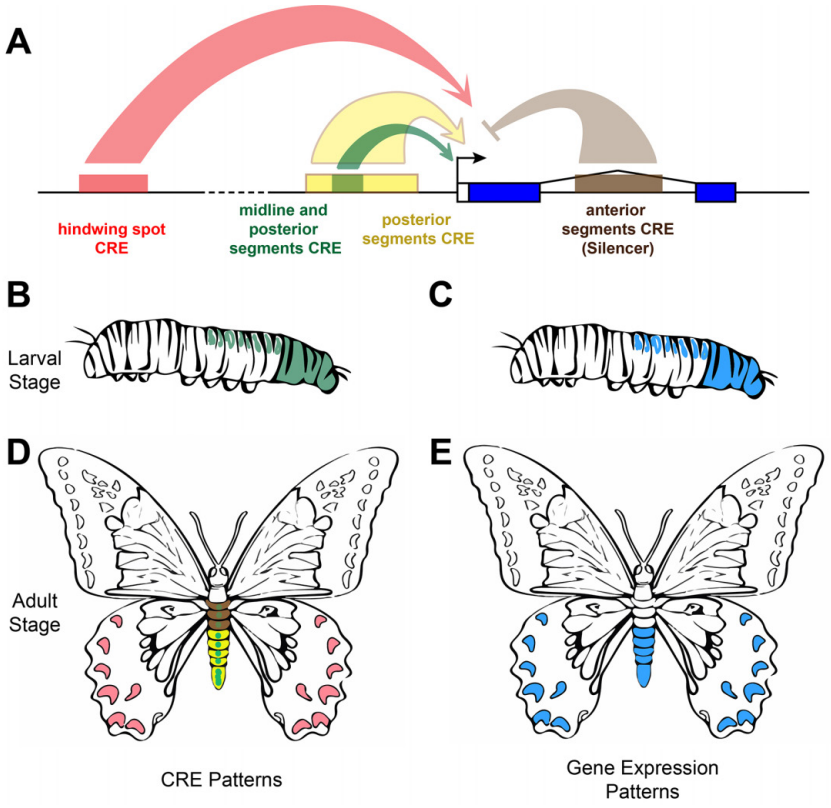

Figure 1.2 The regulatory mechanism of a pleiotropic gene, and its corresponding discrete CRE patterns and gene expression patterns. (A). Schematic showing the regulatory mechanism of a hypothetical pleiotropic gene (blue color) in the butterfly, regulated by multiple discrete CREs. The dashed line indicates that the butterfly hindwing spot CRE (red) acts over a long range to reach the transcription start site. During larval development, the green “midline and posterior segments CRE” is active (B) to drive larval expression (C). As the butterfly metamorphoses into an adult, more CREs become active to facilitate or repress gene expression. Besides the green “midline and posterior segments CRE”, the yellow “posterior segments CRE” (A, D) collaborate to enhance expression in posterior body segments (E). The brown “anterior segments CRE” acts as a silencer in the anterior segments, sculpting the expression to the posterior segments. This gene also plays a role in forming eyespots during late developmental stages (D, E), controlled by a distant eyespot CRE (red). (Please note that the comprehensive regulatory mechanism of butterfly segment specification and eyespot generation has not been fully studied. This is merely a model to show how CREs work cooperatively to regulate the various expression patterns of a pleiotropic gene.)

The functional linkages between the transcription factors and CREs establish gene regulatory networks (GRNs) (**Figure 1.3A**) (Carroll, 2008; Davidson, 2006). In brief, if a gene encodes a transcription factor, such as the dark blue gene in **Figure 1.3A**, it will bind to its downstream targeted gene's CRE (**Figure 1.3A,** yellow, purple and brown genes) to activate or repress transcription (Davidson, 2006). Transcription factors usually have multiple downstream target genes, and function in multiple networks active in different tissues. Their deployment to different tissues is often controlled by multiple CREs which are bound by different transcription factors (**Figure 1.3A,** pink and orange genes). Some gene products even interact with their upstream factors to form feedback loops. Network hierarchies terminate in genes that are generally thought to not regulate other genes in the network. These terminal nodes represent structural genes that execute specific characteristics of the expressing cells, such as shape, position, pigmentation, contractile, elastic, or migratory properties to name a few.

There are two types of genetic alterations that can contribute to a gene expression pattern change, which are changes in *cis* and changes in *trans* (Wittkopp and Kalay, 2011). These refer to the location of mutations in a GRN relative to the gene whose expression is altered. Changes in *cis* occur within the gene itself, altering one of its CREs (**Figure 1.3B**). A *trans*-change involves some upstream factor whose alteration is capable of generating the change in the downstream gene's expression (**Figure 1.3C**, the green gene), while leaving the downstream gene intact (**Figure 1.3C**, the brown and pale blue genes). *trans* changes are likely to affect all of the genes downstream of that factor. It is important to note that an upstream factor responsible for a *trans* change may do so through *cis* regulatory changes, or changes in its protein coding sequence. Finally, it is possible that a gene expression difference is due to a combination of *cis* and *trans* changes that together generate the full effect. There are some ways to estimate the relative contribution of *cis* and *trans* changes to the overall gene expression or phenotypic difference,

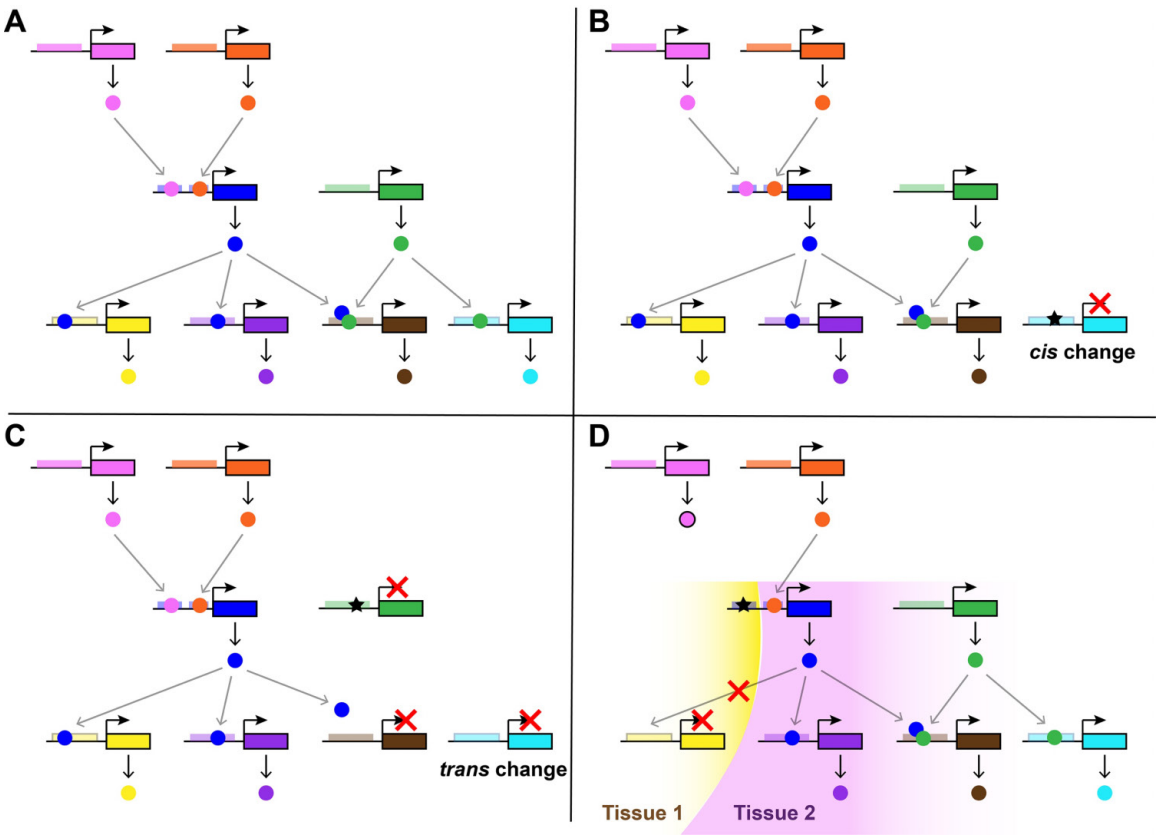

Figure 1.3 Gene regulatory network (GRN) structure and possible evolutionarily relevant alterations in a GRN. (A) GRN architecture. The CREs are represented by narrow rectangles near the transcription start sites, above the black lines. Gene products are represented by round shapes. Pink and orange genes are top-tiers. They both control a same downstream blue gene, through binding to its discrete CREs. The green gene is another high-level factor that affects brown and light blue genes. (B) Schematic showing a *cis*-regulatory change in the light blue-shaded gene—a terminal node in a GRN. The black star indicates the CRE mutation. (C) Schematic showing a *trans*-regulatory change to the brown and light blue genes, caused by the loss of the light green *trans* factor. (D) Schematic showing how the alteration of just one CRE of a pleiotropic gene (the blue gene) can circumvent the pleiotropic effects of its modification during evolution. Because the CRE is active in Tissue 1 (yellow shading), this change only affects expression in Tissue 1, leaving the network of Tissue 2 intact. Black stars indicate the CRE mutations at different nodes.

such as RNA-seq based allele specific measurements of expression in hybrids as well as dissection of QTL to candidate causative loci coupled with analysis of the expression of altered genes and their targets.

Because *trans*-regulatory changes have drastic effects on a GRN, it might appear as though they are too pleiotropic to be a major source of variation during morphological evolution. However, life finds a way to circumvent this pleiotropy and accomplish morphological evolution by altering discrete CREs of pleiotropic genes. In **Figure 1.3D**, the dark blue pleiotropic gene has two discrete CREs which govern its activation in two different networks, Tissue 1 (yellow shading) and Tissue 2 (pink shading). A change that alters the CRE active in Tissue 1 will affect its downstream yellow gene, while other genes active in Tissue 2 remain unaffected. Thus, the modular architecture of gene regulation provides ways to slightly modify a gene active in multiple tissues, even if its alteration has effects on numerous downstream processes.

Next, I will illustrate case studies that have directly implicated *cis*-regulatory changes in the evolution of morphological differences.

1.1.3 Experimentally implicating *cis* regulatory changes underlying phenotypic differences

Multiple approaches exist to directly confirm CRE changes, their effects on expression shifts and ultimately their impact on phenotypic differences. These approaches differ especially on the scale of evolutionary comparison being made. At the population level, deep sequencing based genome wide association (GWAS) studies can associate single nucleotide polymorphisms with differences in phenotype. Genomic regions underlying differences within populations or between species that can be crossed experimentally can be identified through quantitative trait locus (QTL) analysis (Falconer and Mackay, 1996). These two methods represent unbiased approaches to identifying the variation underlying phenotypic differences.

A powerful approach to identify loci underlying traits segregating within populations or between crossable species is the reciprocal hemizygosity test (RHT) (Stern, 2014). The reciprocal hemizygosity test requires that identical mutations, usually null mutations are introduced into the two strains that differ in phenotype. This task has become much easier in recent years, due to the advent of CRISPR/Cas9 genome editing techniques (Bassett et al., 2013; Gratz et al., 2014). In this assay, hybrids of the two species or strains are generated that bear mutations from just one parent. The phenotype of animals containing a mutation from one parent are compared to hybrids that are mutant for the second parent's copy. The end result is that these two groups of hybrids have an identical genetic background throughout the genome, except for the locus of interest. If the phenotypes of these reciprocal hemizygotes differ at all, one can attribute that phenotype to the causative gene that was tested.

However, the use of these methods is limited to populations (GWAS) or species that can be crossed (QTL and RHT). Studies above the species level, in which crosses or GWAS cannot be performed often employ transgenic manipulations to measure the impact of CRE mutations on gene regulation and phenotypes (Kalay and Wittkopp, 2010; Koshikawa et al., 2015; Kvon et al., 2016). In transgenic reporter assays, a specific CRE is cloned from the species that differ in expression phenotype, which are then linked to a fluorescent protein or other reporter gene, and transgenically inserted into the genome of a test strain. The activities of the CREs are then compared in an identical genetic background. Any significant difference of activities would indicate the contribution of that CRE change to the expression shift (Groth et al., 2004; Hosemann et al., 2004; Soriano, 1999). To confirm that CRE changes which affect a gene's expression are phenotypically relevant, these experiments are often complemented by studies in which a rescue transgene containing the implicated change as well as the protein coding unit of the gene are tested side-by-side with a control construct that lacks these changes for differences in phenotypic rescue.

1.1.3.1 *Cis*-regulatory changes cause morphological evolution among populations or crossable species. Numerous examples directly demonstrate the contribution of *cis*-regulatory changes to morphological evolution within populations or between crossable species.

Drosophila usually have different levels of abdominal pigmentation, even within species. Across the continent of Africa, the existence of a striking correlation between abdominal pigmentation and altitude was noted, suggesting that this trait was subject to adaptive evolution (Pool and Aquadro, 2007). This trait was associated with a selective sweep at the yellow-color promoting gene *ebony*, the mutants of which are highly pigmented (Pool and Aquadro, 2007). Analyses of *ebony* expression revealed that indeed differences exist among light and dark strains, and transgenic reporter assays were used to identify five nucleotide substitutions in an abdominal CRE that accounted for the dark phenotype. Transgenic rescue constructs comparing light and dark *ebony* genes confirmed the phenotypic effects of the differences in *ebony* expression (Rebeiz et al., 2009a).

A variety of traits in stickleback fish offer another representative model to study the genes underlying phenotypic shifts. The pelvic spines, which are homologous to the hindlimb of tetrapods, is composed of a pair of bilateral spine bones that connect with an underlying girdle, which protects the sticklebacks from predators (D. Hoogland et al., 1956; Reimchen, 1983). Shapiro et al. performed QTL mapping between marine and freshwater sticklebacks. In the major QTL peak, they identified the *pitx1* gene, which encodes a transcription factor important for hindlimb development (Shapiro et al., 2004). The marine sticklebacks with intact pelvic structures express *pitx1* in the precursor of the pelvic region during larval stages, while the expression was not detected in the freshwater sticklebacks that lack these structures (Shapiro et al., 2004). Later, Chan et al. finely mapped the pelvic regulatory region of *pitx1*, and it was discovered that the pelvic girdle CRE had been deleted in freshwater populations. A transgenic complementation experiment showed that transforming the marine *pitx1* gene into a freshwater strain was sufficient to restore pelvic spines *in vivo* (Chan et al., 2010). A second freshwater trait in sticklebacks is

the loss of armor plating. In these populations, it was noted that *cis*-changes in *Growth/Differentiation Factor 6* (*GDF6*), a member in BMP family of signaling pathway ligands could account for the reduction armor-plate size. Fine mapping and transgenic reporter assay suggested that the freshwater CRE has evolved a transposon insertion which accounts for the higher expression of *GDF6* and smaller armor plates in these populations (Indjeian et al., 2016).

While transgenic complementation assays provide excellent evidence for the phenotypic relevance of an identified change, these assays are not without caveats. In many organisms, targeted transgenesis which inserts transgenes into an identical site each time is not possible. In these cases, such as the stickleback case above, the transgenes are inserted at random locations. This may draw the researcher's attention to the most dramatic insertions. Further the range of phenotypes observed in random transgenesis may introduce more variation than the actual phenotypic difference that is to be measured. Even when inserted into an identical genomic position, all transgenes are taken out of their endogenous context.

The reciprocal hemizygosity test holds exceptional promise in overcoming these caveats of traditional transgenic complementation assays. The power of this approach was recently demonstrated in the case of courtship song differences between *Drosophila simulans* and (*D.*) *mauritiana* (Ding et al., 2016). *Drosophila* males generate "wing songs" during courtship to attract females by vibrating their wings (Ewing and Bennet-Clark, 1968). *Drosophila D*. *mauritiana* has a higher frequency than *D. simulans* in one of the wing song patterns—sine song. Since *D*. *simulans* and *D. mauritiana* just diverged around 0.24 million years ago (Garrigan et al., 2012) and can form viable hybrid offspring capable of generating mating songs, many genetic approaches could be applied to locate and confirm the causative change(s). After fine-scale mapping, Ding et al. located a ~1kb interval showing correlations to the sine song frequency difference between the two species. That interval is located within one of the introns of the calcium-activated potassium channel gene *slowpoke*. The authors then confirmed the role of *slowpoke* as a causal locus underlying sine song frequency differences between the two species through reciprocal

hemizygosity tests (Stern, 2014). To perform the test, null mutations were introduced into the *D. mauritiana* or *D. simulans* copy of slowpoke alternatively in two groups (**Figure 1.4**). The sine song frequency difference between the two hybrids confirmed that *slowpoke* is the causative locus. The authors further dissected the locus through sequencing among populations and genome editing to narrow down the causative mutation to an insertion of a retroelement in the intron in *D. simulans*. This example highlights how mapping, coupled to reciprocal hemizygosity tests, offers an extremely powerful approach to unequivocally identify the causative variation underlying a phenotypic difference.

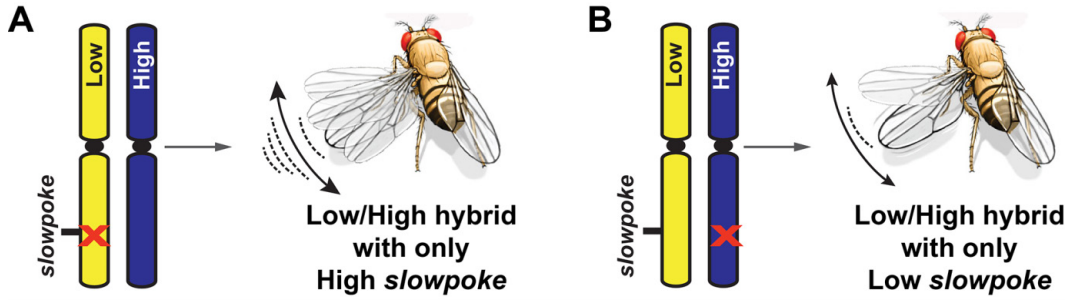

Figure 1.4 Reciprocal hemizygosity tests confirm the causative role for *slowpoke* in generating different wing song frequencies between two closely related species. In these two groups of hybrids, the yellow chromosome is from *D. simulans*, which has a low frequency wing song pattern, while the blue chromosome is from *D. mauritiana*, which has a high frequency wing song pattern. Hemizygotes in (A) bear a functional *D. mauritiana slowpoke* allele, while (B) has a wild type *D. simulans* allele. Any difference in frequency between the two classes of hybrid could reveal a causative role of *slowpoke*.

1.1.3.2 *Cis*-regulatory changes cause morphological evolution between uncrossable species. Unlike *Drosophila* and stickleback fish, uncovering the molecular mechanisms of morphological evolution in distantly related species, such as vertebrates, is far more difficult, but achievements are on their way. A growing number of whole genome sequences have been

obtained from non-model organisms, which allows for genome-wide comparisons of CRE sequences among distantly related taxa and for localizing evolutionarily relevant mutations (Kvon et al., 2016). Even for primates, transcriptome profiling of *in vitro* reprogrammed and differentiated cells also provided clues of CRE alterations that are associated with human-chimpanzee craniolfacial evolution (Prescott et al., 2015). To confirm the differential activities of these CREs in vivo, transgenic reporter assays in mouse were applied to compare these candidate causative alterations in CREs for these traits (Kvon et al., 2016; Prescott et al., 2015), providing evidence of changes underlying the evolution of traits that would be impossible to confirm otherwise.

1.2 HOX GENES AND THE MODIFICATION OF ANIMAL BODY PLANS

1.2.1 Hox gene expression patterns correlate with shifts in animal body plans on a macroevolutionary scale.

The examples in the above section demonstrate how CRE alterations often drive phenotypic evolution. It is worth noting that each of the case studies which confirmed the phenotypic contribution of CRE changes were performed among populations, i.e., on a microevolutionary scale, or between extremely closely related species that can be crossed experimentally. However, in the animal kingdom, the morphological variations that we would like to explain often exist among distantly related taxa that are not crossable. These might include the elongated axial skeleton of snakes, the striking diversity of appendage types in crustaceans, the exaggerated legs of water striders, or the loss of hindwings in dipteran flies, to name a few. All of these differences occur in the animal's body plan, which refers to the number and specialization of serially repeated parts (e.g. serial homologs) such as limbs or vertebrae (Valentine, 2004). Body plan changes typically differ among taxa above the species level, and are thus macroevolutionary

(Dobzhansky, 1937; Reznick and Ricklefs, 2009). For decades, evo-devo biologists have sought to uncover the genetic basis, especially the major driving force that accounted for such shifts in the body plan.

A group of transcription factors, the Hox genes have long been known to pattern and organize the body plan among metazoans. Expression of the Hox gene *Ultrabithorax* (*Ubx*), known to be important in thoracic appendage modification in *Drosophila* (Lewis, 1964; Morata and Garcia-Bellido, 1976), is patterned in concert with the varying numbers of feeding legs across crustacean species (Averof and Patel, 1997). In crustacean species that lack feeding legs such as *Artemia* (**Figure 1.5A**), the Hox genes *Ultrabithorax* (*Ubx*) and *abdominal-A* (*abdA*) are expressed throughout the thoracic segments (**Figure 1.5C**). In contrast, *Periclimenes* has evolved a more derived body plan, which includes three pairs of feeding legs in its first three thoracic segments (**Figure 1.5B-C**). Accordingly, Ubx-AbdA expression is reduced in the segments that produce these modified appendages, and begins only at the fourth thoracic segment, which bears unmodified locomotory appendages (**Figure 1.5C**).

This example illustrates how Hox gene expression changes are correlated with major phenotypic shifts, such as altering the specialization of segmentally repeated appendages. Many examples in the evo-devo field show similar correlations between Hox genes and body plan shifts between distantly related taxa, including vertebrates, insects, spiders, as well as centipedes (Cohn and Tickle, 1999; Damen et al., 1998; Hughes and Kaufman, 2002; Khila et al., 2009; Stern, 1998). Since Hox genes play such a critical role in affecting body plan formation (McGinnis and Krumlauf, 1992) it is important to consider their structured regulatory mechanisms in detail.

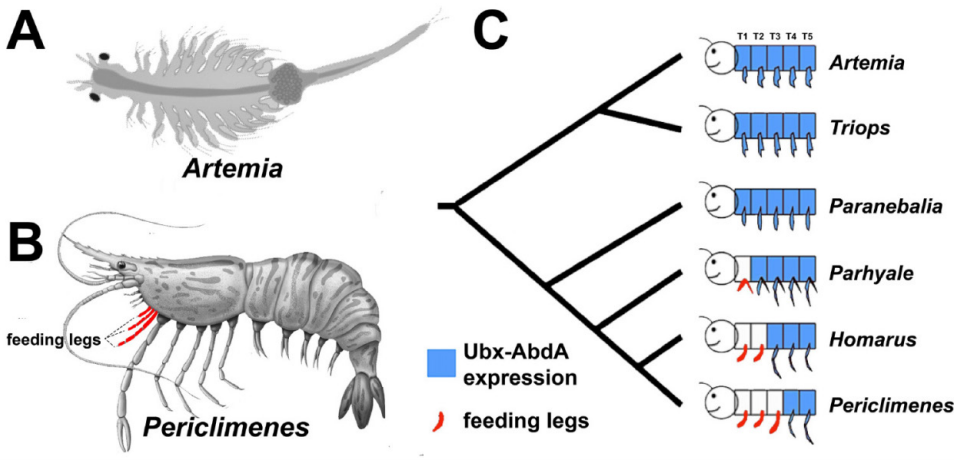

Figure 1.5 Hox gene expression changes track with morphological differences in repeated body parts among crustacean species. (A). *Artemia* brine shrimp possess an ancestral body plan that lacks feeding legs. The appendages along the thorax are swimming legs. (B). A more highly specialized crustacean species *Periclimenes* possesses three pairs of feeding legs, colored red. (C). Phylogenetic tree showing the distribution of appendages and their respective expression patterns of the Hox genes Ubx-AbdA (blue shading). T1 to T5 indicate the first five thoracic segments. Red appendages indicate feeding legs. Panel C adapted from (Averof and Patel, 1997).

1.2.2 The identification of Hox gene clusters and their regulatory mechanisms.

The identification and elucidation of *cis*-regulatory mechanisms of Hox genes were first unraveled and by Lewis (Lewis, 1978), for which he was awarded the Nobel Prize in 1995. Lewis noted a class of mutations that generated homeotic transformations (transforming one body segment's identity into that of another) which were organized in a cluster on the chromosome. Intriguingly, their relative positions on the chromosome were co-linear with the corresponding body segments that were affected (Lewis, 1978). This led to the identification of the bithorax complex (BX-C), a ~300 kb genomic region containing three Hox genes, *Ubx*, *abd-A* and *Abdominal-B* (*Abd-B*).

The Hox genes were found to encode transcription factors which share a conserved helix-turn-helix motif called homeodomain that recognizes and binds to a specific DNA sequence (Gehring et al., 1994). However, the sequence they recognize, TNATNN, in which N is variable, occurs at high frequency throughout the genome (Ekker et al., 1994; Merabet et al., 2010). The N- and C-terminal sequences of the homeodomain contribute to functional specificity of a portion of *Drosophila* Hox proteins (Chan and Mann, 1993; Chauvet et al., 2000; Furukubo-Tokunaga et al., 1993; Lin and McGinnis, 1992; Zeng et al., 1993), and cofactors such as *Extradenticle* (*Exd*), can modulate binding specificity, especially in the case of low affinity binding sites (Crocker et al., 2015). However, the coding sequences and exon/intron architectures of Hox genes possess a high level of conservation among metazoans (McGinnis et al., 1984; Yoder and Carroll, 2006). Furthermore, these genes are highly pleiotropic— they start to function from early embryonic stages (Carroll, 1995; Hirth et al., 1998; Manak and Scott, 1994; McGinnis and Krumlauf, 1992; Van Auken et al., 2000), are expressed in multiple tissues (Averof and Patel, 1997; Carroll, 1995; Cohn and Tickle, 1999), and act as high level factors in gene regulatory networks (Hu and Castelli-Gair, 1999; Kopp et al., 2000; Parker et al., 2018; Williams et al., 2008). Thus, a key to their role in evolution is very likely to be related to how they are regulated. Fortunately, their CREs have been subject to intense study.

Analysis of specific alleles contained within the BX-C identified CREs that drive expression of *Ubx*, *abd-A,* and *Abd-B* (Celniker et al., 1990). Deficiencies of the entire BX-C generated offspring that had drastic morphological transformations of all abdominal segments into anterior thoracic segments, and yet smaller local deletions within the complex had more specific effects on one or a few segments (Lewis, 1978). In a fine-scale characterization of the regulatory region between *abd-A* and *Abd-B*, mutations in the *iab-6* region were found to reduce *Abd-B* expression in the A6 segment, resulting in its transformation to the A5 segment (Celniker et al., 1990). Deletion of *iab-5* resulted in the loss of Abd-B expression from the A5 segment, leading to the transformation from A5 to A4 (**Figure 1.6B**) (Celniker et al., 1990).

Mutations within the BX-C also revealed insulator elements that compartmentalize the activity of these CREs. For example, *iab-5* is normally silenced in A4 by the insulator contained in the *Miscadestral pigmentation* (*Mcp*) region (Busturia et al., 2001; Lewis, 1978). *Mcp* sets up a boundary between *iab-4* and *iab-5* (Karch et al., 1994). Deficiencies that remove *Mcp* allow *iab-5* to drive activity in the A4 segment, causing an A4 to A5 transformation (**Figure 1.6D**) (Celniker et al., 1990).

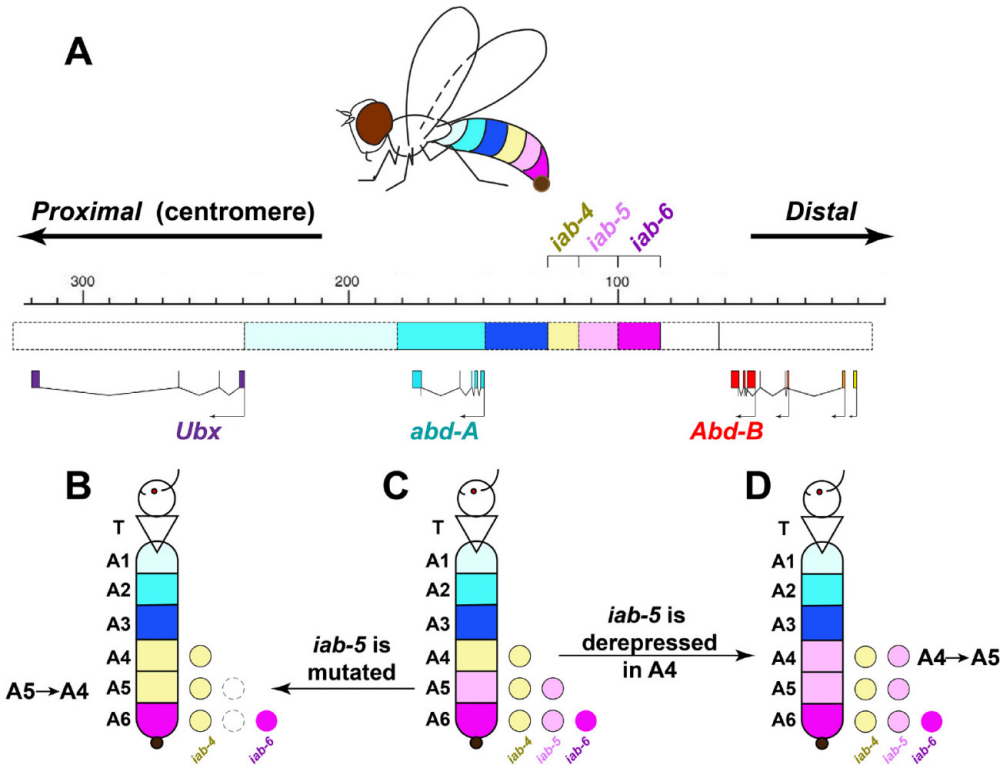

Figure 1.6 Regulatory elements of the BX-C and their associated homeotic transformations. (A). Diagram of the ~300kb BX-C showing the relative positions of the three Hox genes *Ubx*, *abd-A*, *Abd-B* and a series of tandem *infra-abdominal* (*iab*) regulatory sequences along the chromosome. Multiple transcription start sites in *Abd-B* depict different transcript isoforms. The abdominal segments of the male fly model are color-coded collinearly with the *iab* regulatory elements organized along the genome. The proximal and distal directions of the BX-C are shown in black arrows. (C). Activity distributions of the *iab* regulatory sequences in wild type male abdominal body segments A1-A6. The thorax is shown as a triangle, marked by "T". (B, D). Demonstrations of homeotic transformations. For ease of presentation, only *iab-4*,*5* and *6* elements are shown. (B) The *iab-5* mutation causes inactivation of *iab-5* in A5 and A6. Thus, A5 has the BX-C expression patterns of A4, leading to the transformation from A5 to an A4 phenotype (Celniker et al., 1990). (D) In the *Mcp* mutant, *iab-5* is not insulated from abdominal segment A4, driving A5 levels of BX-C gene activity in the A4 segment, resulting in the transformation from an A4 to an A5 phenotype. Panel A adapted from (Maeda and Karch, 2006). Panels B-C adapted from (Lewis, 1978).

1.2.3 Challenges in implicating Hox CREs during body plan evolution

Hox genes are thus very important and highly conserved regulators of animal body plans. Differences in their expression patterns often correlate with evolutionary changes in the body plan among distantly related species. Considering the conservation of Hox gene coding sequences, their clustered organization in metazoan genomes, as well as their high level of pleiotropy, it would seem likely that their CREs would be a frequently traversed path for evolutionarily relevant changes. Work in the *Drosophila* Hox complexes have identified complex arrays of CREs that govern the precise temporal and spatial expression of Hox genes during development (**Figure 1.6**). Therefore, one might expect that a large number of examples would have quantified their contributions to differences in animal body plans, and pinpointed their causative changes.

However, it has been quite challenging to directly implicate Hox genes in phenotypic traits and localize their causative changes. Importantly, in macroevolutionary cases, where body plan shifts are most obvious, such as the crustacean example (**Figure 1.5**), comparisons involve taxa which are separated by hundreds of millions of years. As a result, such differences cannot be dissected by quantitative methods that involve crosses or complementation tests, such as QTL analysis. Such long divergence time also results in very low sequence identities among species, especially for the rapidly evolved non-coding CRE sequences (Rebeiz et al., 2015), making it difficult to identify a small number of causative changes when they are vastly outnumbered by other non-relevant variation. The tests that one might perform on such regulatory regions to examine potential CRE differences involve transgenesis which is typically hard to perform in non-model species, and is complicated by technical issues when comparing activities among distantly related taxa (Arnosti, 2003).

A second major problem with pinpointing the contributions of Hox genes to body plan changes is that these genes cannot function in isolation. Hox genes have such drastic effects on phenotypes because they govern a gene regulatory network where other genes also play roles.

While evolutionary changes may occur in the Hox gene to alter the phenotype, they could just as easily reside in their upstream factors or downstream targets to alter the output of the network. Indeed, one attempt to reconcile how Hox genes participate in the evolution of body plans suggested that the networks surrounding Hox genes evolve first, and their evolutionary divergence arises only after the phenotype has changes as a stabilizing effect (Budd, 1999). Thus, it's a challenge to measure the relative contribution of Hox genes to trait evolution in a non-biased way. Therefore, we need a model system where we can study Hox-regulated body plan evolution among closely related species in which genetic analyses can identify and measure the relative contribution of each member to the phenotype, and unravel the positions and interactions among contributing genes in a gene regulatory network.

1.3 *DROSOPHILA* ABDOMINAL PIGMENTATION—A PREMIERE MODEL TO STUDY THE GENETICS UNDERLYING HOX-REGULATED BODY PLAN EVOLUTION

1.3.1 *Drosophila* abdominal pigmentation—a Hox-regulated polygenic trait

The rapidly evolving abdominal pigmentation patterns of *Drosophila* species have been a particularly fruitful platform in which to study the nature of developmental evolution (Rebeiz et al., 2009a; Wittkopp et al., 2003; Wittkopp et al., 2002b). Despite ample phenotypic divergence between species, the metabolic pathway underlying pigmentation formation is highly conserved (Wittkopp et al., 2003). Each abdominal segment forms an epithelium whose anterior compartment secretes a rigid cuticle that exhibits different patterns of pigmentation in different segments (Kopp and Duncan, 1997; Struhl et al., 1997). The process of generating pigmentation occurs in abdominal epidermal cells where the precursors of pigments are secreted out to develop into black melanin, brown melanin, or yellow sclerotin, catalyzed by phenol oxidases (**Figure 1.7**) (Wittkopp et al., 2002a; Wright, 1987). The pigments are all derived from DOPA and dopamine,

which are synthesized from tyrosine. DOPA decarboxylase (*Ddc*) converts DOPA to dopamine, the precursor of brown melanin. While multiple enzymes in the pathway have been characterized biochemically, there are three structural genes—*yellow*, *tan* and *ebony*— which are highly patterned during the development of the abdominal cuticle (**Figure 1.7**). When the *yellow* gene is active, DOPA is converted to black melanin. Although the exact enzymatic function of the Yellow protein is unknown, a recent study reported that it is responsible for forming the sharp boundaries of black pigments, and that it likely anchors forming pigments or other pigmentation enzymes in expressing tissues (Hinaux et al., 2018). Notably, experimental manipulation of *yellow* homologs across a wide variety of taxa shows defects in black melanin formation (Arakane et al., 2009; Futahashi et al., 2008; Liu et al., 2016), indicating a deeply conserved role for this protein's function. In the presence of *ebony* expression, dopamine is converted to N-β-alanyl-dopamine (NBAD), which is used to produce yellow sclerotin (Wittkopp et al., 2002a). NBAD can be converted back to dopamine by *tan,* which encodes an NBAD hydrolase enzyme (True et al., 2005; Wright, 1987). Therefore, *yellow* and *tan* genes promote dark pigment formation, while *ebony* drives yellow color patterns.

To generate sharp patterns of pigmentation, *yellow*, *ebony* and *tan* are precisely regulated during abdominal development (Rebeiz et al., 2009a; Walter et al., 1991; Wittkopp et al., 2002a). The activation of these terminal network genes is controlled by their upstream transcription factors in the pigmentation gene regulatory network. The pattern of *Drosophila melanogaster* abdominal pigmentation is controlled, ultimately, by the activity of Hox genes, such as *Abd-B,* required for sexually dimorphic pigmentation of the fifth and sixth body segments of males (Celniker et al., 1990). In 2006, Jeong et al. identified ABD-B binding sites in the abdominal CRE of *yellow*, which are required for the activity of this CRE in the abdomen (Jeong et al., 2006). The *tan* gene, which is expressed in a pattern similar to *yellow* is indirectly regulated by *Abd-B* and directly repressed by the Hox gene *abd-A* in more anterior segments (Camino et al., 2015). Since the regulatory mechanism of Hox genes have been intensively studied in *Drosophila* (Celniker et al., 1990; Lewis,

1978; Maeda and Karch, 2006; Mihaly et al., 2006), the rapidly evolving pigmentation patterns on the abdomens of *Drosophila* species closely related to *D. melanogaster* offer an ideal model system for studying how they contribute to phenotypic differences in body plan networks.

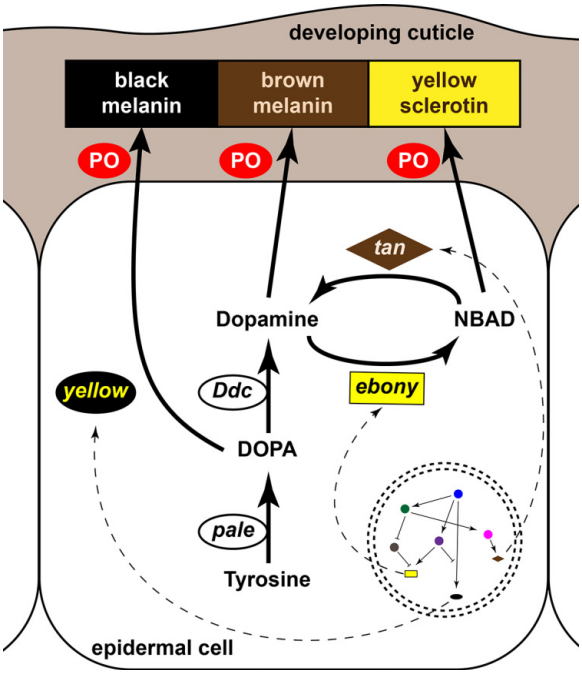

Figure 1.7 The *Drosophila* abdominal pigmentation pathway, derived from (True et al., 2005). The figure shows three of the parallel metabolic pathways where DOPA and dopamine are converted to the precursors of colored pigment deposits and secreted into developing cuticles to be catalyzed by phenol oxidase (PO, red ovals). *yellow*, *tan* and *ebony* are structural genes that play critical roles in patterning pigment formation in the cuticle. The nucleus of the epidermal cell (double dashed-line circle) depicts a gene regulatory network that governs the expression of these genes.

1.3.2 The recently diverged sister species *D. yakuba* and *D. santomea* represent a genetically tractable model for pigment evolution

Among the nine members of the *melanogaster* subgroup, which contains the closest relatives of *D. melanogaster*, *D. santomea* was discovered on the island habitat of São Tomé Island (Lachaise et al., 2000), where it diverged from its sister species, *D. yakuba*, around 0.5 to 1 million years ago (Bachtrog et al., 2006; Obbard et al., 2012). In spite of their recent divergence, male *D. santomea* exhibit pale yellow abdomens (**Figure 2.1B**) compared to *D. yakuba* (**Figure 2.1A**) and other *melanogaster* subgroup species with very dark pigmentation in the posterior two body segments (A5-A6). These two species diverged from *D. melanogaster* approximately 5-10 million years ago (Bachtrog et al., 2006; Obbard et al., 2012). This small window of evolutionary time guarantees that most of the molecular mechanisms of pigmentation formation discovered from *D. melanogaster* likely also be applied to *D. yakuba* and may be modified in *D. santomea*, such as the regulatory mechanisms of BX-C (Celniker et al., 1990; Maeda and Karch, 2006).

Previously, a quantitative trait locus (QTL) study of pigmentation differences between these species were mapped to four separate genomic regions: one strong and one very weak QTL on opposite ends of the X chromosome, and one QTL each on the second and third chromosomes (Carbone et al., 2005). The strong QTL on the X chromosome is centered over the candidate gene *tan*, and functional and mapping studies strongly implicated variation at the *tan* gene as the cause of this QTL (Jeong et al., 2008; Rebeiz et al., 2009b). The other genes underlying the other QTLs were unknown at the beginning of my thesis research (Jeong et al., 2008).

In my thesis, I will illustrate how I identified four additional genes that evolved to generate loss of pigmentation *in D. santomea* and confirmed their contributions to the phenotypic difference. All of the five identified causal genes (*Abd-B*, *pdm3*, *ebony*, *yellow*, and *tan*) fall within the previously published QTL peaks. They act together in a genetic network regulated, at least

partially, by the Hox gene *Abd-B* and the POU domain transcription factor *pdm3*. I have confirmed that all five genes have evolved through changes in their CREs. Each altered gene has a relatively strong effect on pigmentation on its own, but in the context of the other evolved genes in the GRN, each gene displays a much smaller and distinct phenotypic effect. My use of genetic approaches to directly implicate the involvement of a body-plan patterning Hox gene provides a key example of a polygenic morphological trait, and offers a link between micro- and macroevolution. Seeing how Hox genes collaborate with their downstream networks to generate phenotypic differences sets an important precedent for anticipating such complexity when small changes are compounded over hundreds of millions of years to generate macroevolutionary differences.

2.0 A NEAR COMPLETE DISSECTION OF A HOX-REGULATED NETWORK THAT DRIVES EVOLUTION BETWEEN TWO SISTER SPECIES

In this chapter, I will present a genetic and molecular dissection of the mechanisms through which a Hox-regulated network evolved to generate a drastic shift in abdominal pigmentation between two sister species of *Drosophila*. By integrating introgression mapping and gene expression analysis, I identified five candidate genes showing evolved expression patterns between *D. yakuba* and *D. santomea*. These genes span the hierarchy of a gene regulatory network, including top-tier transcription factors and terminal enzyme encoding genes. Using transgenic reporter assays, I narrowed down the *cis* regulatory mutations in most of these loci, including the Hox gene *Abd-B*. I then confirmed the contribution of each locus to the phenotypic difference by performing reciprocal hemizygosity assays. My results demonstrate that a Hox-regulated network has evolved at multiple tiers, illustrating for the first time how microevolutionary changes at multiple genes in a single regulatory network generated a large morphological difference between species.

Please note that in this Chapter, the introgression lines, MSG genomic maps and the CRISPR mutant flies were generated by our collaborator, Dr. David Stern. Examination of Ubx expression was carried out by Dr. William Rogers in the lab of Dr. Thomas Williams. The population surveys of *D. yakuba* and *D. santomea* isofemale lines were performed by Clair Han, a graduate student in the Andolfatto Lab.

2.1 COORDINATED EVOLUTION OF THE EXPRESSION OF *yellow* AND *ebony* EXPRESSION IN *D. SANTOMEA*

A general trend of pigment pattern evolution is that multiple pigment producing enzymes exhibit coordinated patterns of gene expression that change in concert. While the black pigment-promoting *tan* and *yellow* genes tend to be co-expressed, *ebony* is deployed in an inverse pattern with respect to these two genes (Camino et al., 2015; Gompel et al., 2005; Johnson et al., 2015; Ordway et al., 2014; Wittkopp et al., 2002a). Previous work had shown that the expression of both *yellow* and *tan* was lost in the posterior abdominal segments of *D. santomea* (Jeong et al., 2008). Given the importance of *ebony* in generating lighter-colored regions, I was curious whether *ebony* underwent a complementary expansion in expression to generate the light-colored abdomen of *D. santomea*. While *D. yakuba* showed the expected expression of *ebony* transcripts in anterior body segments (**Figure 2.1K**), its expression in *D. santomea* was expanded posteriorly (**Figure 2.1L**). Thus, expression of these genes is consistent with the observed pigmentation differences between the two species, which could result from changes in the CREs of *yellow* and *ebony* and/or changes in *trans* regulators of these genes. This result led me to examine the root genetic causes of how *ebony* was gained, and *yellow* was lost in the *D. santomea* abdomen.

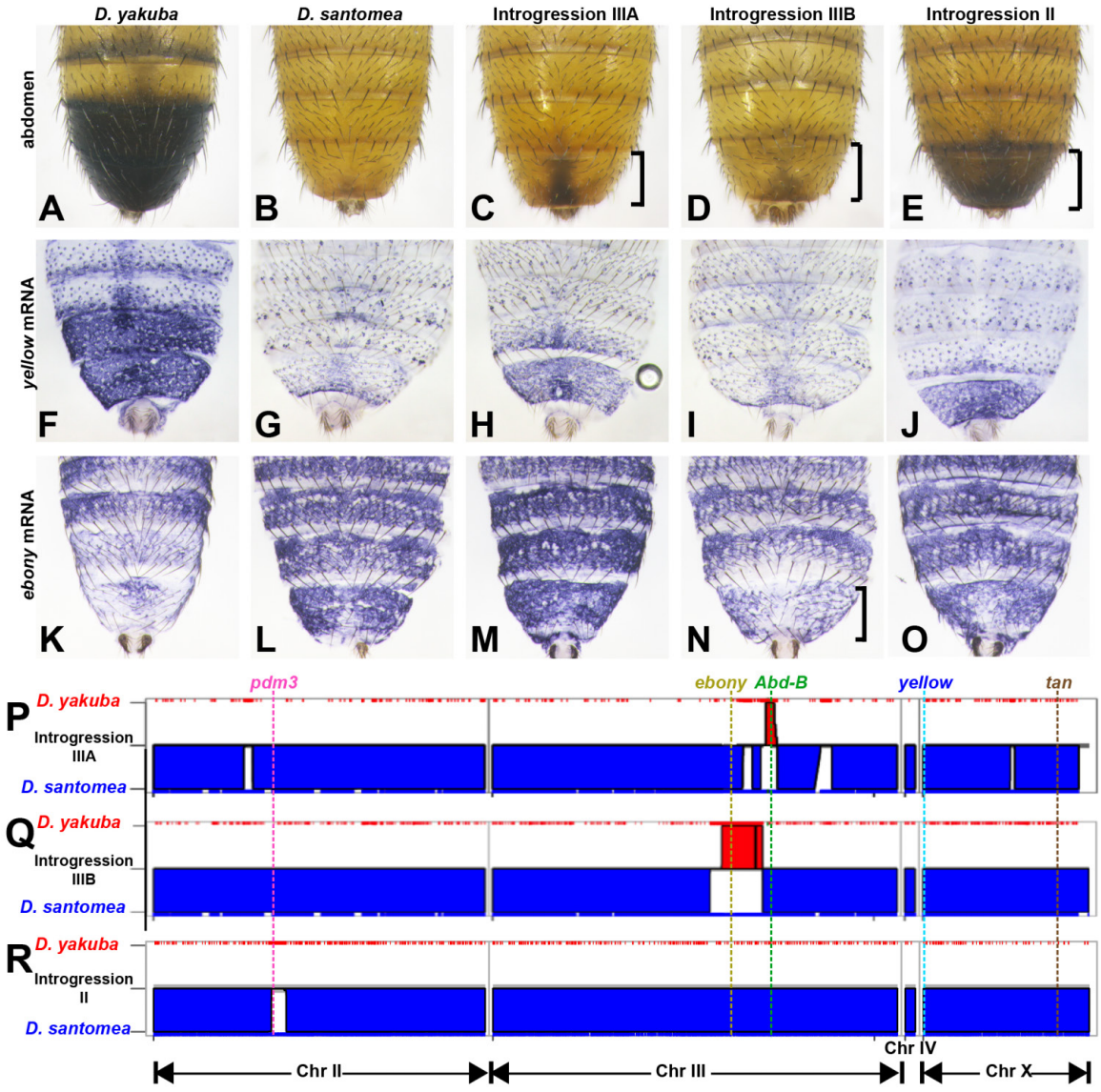

Figure 2.1. Distinct genetic causes underlying coevolution of *ebony* and *yellow* expression in *D. santomea*. (A-E) Male abdominal pigmentation patterns in the posterior body segments. (A) *D. yakuba* body segments A5 and A6 are fully pigmented. (B) *D. santomea* lost male-specific pigmentation in A5 and A6. (C) The Introgression IIIA line has black pigmentation in the midline of the A6 segment (black bracket). (D) The Introgression IIIB line exhibits increased pigment in A6 (black bracket). (E) The Introgression II line greatly expands its black pigments in A6. (F-J) Visualization of *yellow* expression by *in situ* hybridization of the male pupal abdomens, ~24 hours preceding eclosion. While *D. yakuba* expresses *yellow* throughout the A5 and A6 segments (F), *D. santomea* lacks expression (G), a phenotype that is also observed in the

Introgression IIIB line (I). However, expression of *yellow* is partially restored in the Introgression IIIA and II lines (H, J). (K-O) *ebony* mRNA accumulation in the male abdomen within 1 hour post-eclosion, revealed by *in situ* hybridization. (K) In *D. yakuba*, expression is greatly reduced in A5 and A6 segments. (L, M) *D. santomea* and Introgression IIIA both exhibit high expression throughout the abdomen, while Introgression IIIB (N, black bracket) has a *D. yakuba*-like pattern, in which expression is greatly reduced in A6. (P-R) Genomic DNA maps of the introgression lines from multiplexed shotgun genotyping (MSG) sequencing of Introgressions IIIA (P), IIIB (Q), and II (R). Four pairs of chromosomes are indicated below the map. Blue regions are homozygous for *D.santomea*, while homozygous D. yakuba regions are shown in red. Regions of the map that are not blue or red indicate heterozygosity. Dashed lines indicate the positions of *tan*, *yellow*, *Abd-B*, *pdm3* and *ebony* genes.

2.2 INTROGRESSION MAPPING UNCOVERS DISTINCT GENETIC CAUSES UNDERLYING EXPRESSION DIFFERENCES OF PIGMENTATION GENES

Traditional QTL mapping involves testing for statistical associations between genetic markers and phenotypic features in a large mapping population. Statistical power is determined largely by the sample size and precision is determined largely by the amount of recombination in the mapping population. In *Drosophila*, QTL mapping usually provides low resolution of causative loci, because the genome contains only four chromosomes, each of which experience only one or two recombination events per generation (Ashburner et al., 2005). My colleague Margarita Ramos-Womack, a member of the Stern lab, attempted to overcome these limitations using phenotype-based selection and introgression (Earley and Jones, 2011). Dr. Ramos-Womack backcrossed recombinant offspring of a cross between *D. yakuba* and *D. santomea* repeatedly to *D. santomea* for seven generations. In each generation, individuals were selected that exhibited more pigmentation on their abdomen than is observed in *D. santomea*. After attempting to generate homozygous lines for these introgressions by sib-mating individuals from the same selection line,

multiplexed shotgun genotyping (MSG), a high-resolution whole-genome genotyping assay (Andolfatto et al., 2011), was performed to map the locations of the introgressions in all lines.

Among these introgressions, I found lines that mapped to regions that correspond with previously identified QTL on the second and third chromosomes (Carbone et al., 2005). On the second chromosome, Introgression II greatly restored pigment to the A6 abdominal segment (tergite), representing the strongest single effect isolated in my introgression analysis (**Figure 2.1E**). I identified two lines that contained partially-overlapping regions of the third chromosome of *D. yakuba*, yet produced distinct patterns of pigmentation (**Figure 2.1C, D**). While Introgression IIIA generated strong midline pigmentation (**Figure 2.1C**), Introgression IIIB subtly darkened the A6 tergite and intensified the pigmented bands forming at the posterior of each tergite (**Figure 2.1D**). The distinct phenotypes of the third chromosome introgressions suggested the existence of causative loci outside of the overlapping region.

To investigate how these introgression lines manifest their distinct phenotypes, I examined their effects on pigmentation enzyme expression. As previous work attributed a major-effect QTL on the X chromosome to regulatory changes that inactivated *tan* expression in *D. santomea* (Jeong et al., 2008; Rebeiz et al., 2009b), I focused the analysis on the other two structural genes *yellow* and *ebony*. Two of our introgressions, Introgression II and IIIA caused increases in *yellow* expression (**Figure 2.1H, J**), but had no effect on *ebony* (**Figure 2.1M, O**). Meanwhile, Introgression IIIB showed reduced expression of *ebony* that inversely correlated with its pigmentation phenotype (**Figure 2.1N**), but had no effect on *yellow* (**Figure 2.1I**). These results indicate that the chromosome II and III QTLs have *trans*-acting effects on *yellow* that caused its loss of expression. Furthermore, the independent effects of third chromosome introgression lines on *yellow* and *ebony* corroborated the presence of distinct causative loci that lie outside their minimally overlapping interval, suggesting that the two introgressions captured different genes

contributing to pigmentation variation. Having characterized the impact of our introgressions on pigmentation enzyme expression, I next sought to identify the genetic differences that cause these phenotypes.

2.3 IDENTIFICATION AND EXAMINATION OF *CIS* REGULATORY CHANGES AT THE HOX GENE *Abd-B* IN *D. SANTOMEA*

The MSG map of the Introgression IIIA line identified five regions of the *D. yakuba* genome present in an otherwise *D. santomea* background (**Figure 2.1P**). The *yellow* gene, whose expression is altered in this line (**Figure 2.1H**), is nevertheless located outside of these regions (**Figure 2.1P**), suggesting that the introgressed regions contain one or more factors upstream of *yellow* that differ in activity between *D. yakuba* and *D. santomea*. Furthermore, only one genomic region was homozygous in the MSG maps (**Figure 2.1P**). This introgressed region contained the *D. yakuba* bithorax complex (BX-C), which is a key candidate locus for several reasons. In *D. melanogaster*, the Hox genes of the BX-C, *Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*) and *Abdominal-B* (*Abd-B*) control the identities of the third thoracic (T3) and first through ninth abdominal (A1-A9) segments (Lewis, 1978). Of these, *Abd-B* specifies identity in the posterior abdominal segments, and its expression is generally conserved among metazoans (Kenyon and Wang, 1991; Malicki et al., 1990; McGinnis et al., 1990). As Abd-B is a direct regulator of *yellow* through binding to a CRE just 5’ of its first exon (Jeong et al., 2006), I examined the possibility that Abd-B expression was altered in *D. santomea*.

2.3.1 The Hox gene *Abd-B* evolved temporally restricted expression changes

Although *Abd-B* is known to be required for posterior abdominal pigmentation, its expression during late pupal development has not been described. Considering its direct regulation of the abdominal CRE of *yellow*, I speculated that the protein must be present during or just preceding the time when *yellow* is expressed in the abdomen. While *D. yakuba* displayed the expected expression pattern in the A5 and A6 tergites of late stage pupae, *D. santomea* showed expression restricted to just the A6 tergite (**Figure 2.2A, B**). To test whether the difference in Abd-B expression between the two species was caused by *cis*-regulatory changes, I tested Abd-B expression in our introgression lines. Although Introgression II and IIIB showed the *D. santomea* pattern of Abd-B expression (**Figure 2.2D, E**), Introgression IIIA, which contains the *D. yakuba* *Abd-B* ortholog, restored the A5 expression phenotype (**Figure 2.2C**).

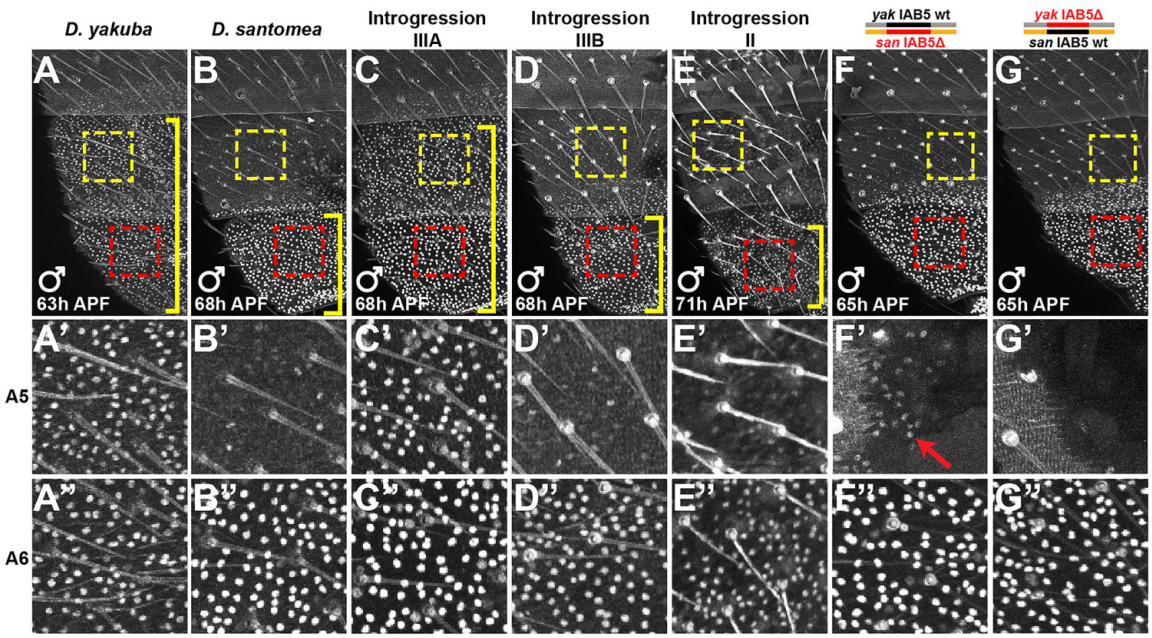

Figure 2.2 The evolution of reduced *Abd-B* expression in the *D. santomea* abdomen. (A-E) Abd-B immunostainings of 63-68 hAPF pupal abdomens. Images show the posterior segments. Yellow brackets indicate the extent of expression. (A) In *D. yakuba*, Abd-B is expressed in the A5 and A6 body segments. (B) *D. santomea* exhibits A6 expression. (C) The Introgression IIIA line shows the *D. yakuba* expression pattern of Abd-B. (D) In Introgression IIIB, which contains the *D. santomea Abd-B* allele, Abd-B expression is limited to the A6 segment. (E) The Introgression II line shows a *D. santomea* expression pattern. (F, G) Abd-B expression in 65 hAPF pupal abdomens of the two groups of progeny from reciprocal hemizygosity tests of the IAB5 deletion. (A'-G', A"-G") Closeup images of the areas outlined in (A-G), showing epithelial cell expression in A5 (yellow box, A'-G') and A6 (red box, A"-G").

Since Hox genes play critical roles in segmental identity specification in early embryonic developmental stages, I was curious whether the observed differences in Abd-B spatial distribution extended across development. In embryos and larvae, Abd-B expression was indistinguishable between *D. yakuba* and *D. santomea* (**Figure 2.3A-D**). Following puparium formation (0 hours after puparium formation or hAPF), the histoblast nests expand to cover the abdomen by 26 hAPF. At this early stage of abdominal development, Abd-B expression is spatially indistinguishable between these two species (**Figure 2.3E, F**). Furthermore, examination of the expression of the two other BX-C genes, Ubx and Abd-A, revealed conserved patterns of expression between *D. yakuba* and *D. santomea* (**Figure 2.3G-J**). These results suggested that changes to the BX-C are limited specifically to *Abd-B*, altering gene expression during a developmental window between 26 and 68 hAPF.

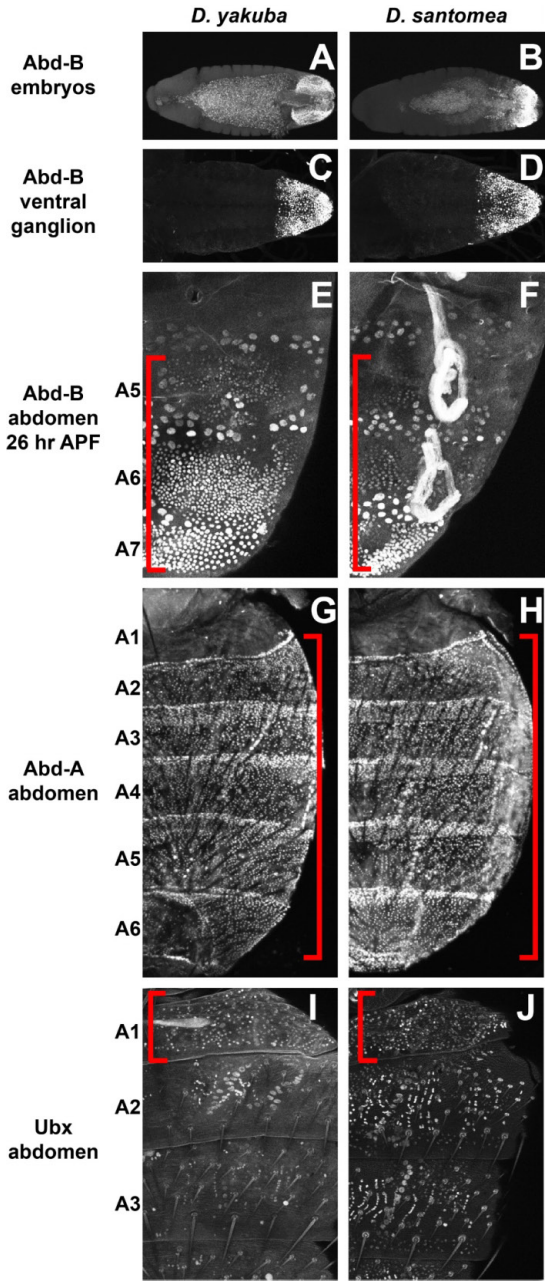

Figure 2.3 Expression analysis of the Hox genes of the bithorax complex. (A-F) Abd-B is expressed similarly in *D. yakuba* and *D. santomea* embryos (A-B), larva (C-D), and early pupal abdomen (E-F). (G-H) Abd-A is expressed similarly between *D. yakuba* and *D. santomea* throughout the mid-pupal abdomen (~68h APF). (I-J) Ubx is limited to epithelial nuclei of the A1 abdominal segment in late pupae of both *D. yakuba* and *D. santomea*. Red brackets highlight the expression patterns in pupal stages.

2.3.2 Changes to the *cis*-regulatory region of *Abd-B* contribute to its expression and phenotype in *D. santomea*.

To determine whether *Abd-B* and other candidate genes in this study contributed to phenotypic differences between *D. yakuba* and *D. santomea*, I performed reciprocal hemizygosity tests (RHT) (Stern, 2014) in this species pair, using CRISPR/Cas9 genome editing to generate targeted genic deletions (Gratz et al., 2014). In this test, equivalent lesions are introduced into a gene suspected to contribute to the phenotypic differences between two strains or crossable species. The mutant of the first strain is crossed to the wild type of the second strain, and vice versa. If the phenotypes of these reciprocal hemizygotes are different, this difference can be confidently attributed to allelic variation in the mutated gene, as it represents the sole source of variation that exists among compared individuals.

As a positive control, I performed a reciprocal hemizygosity test using null alleles of *tan*, which has been implicated previously as the causative gene underlying the large QTL on the X chromosome (Jeong et al., 2008; Rebeiz et al., 2009b). Using CRISPR/Cas9 gene editing, null alleles of *tan* that removed exons encoding its catalytic domain were generated and replaced by a *3xP3*::*DsRed* reporter (Horn et al., 2000), which drives a fluorescence marker in the eyes. Flies homozygous for these mutant alleles show reduced pigmentation in females (**Figure 2.4C, D**), confirming the conserved role of tan in abdominal pigmentation. Male phenotypes are less obvious (**Figure 2.4A, B**) due to the strongly patterned expression of other pigmentation genes in this background (**Figure 2.1F**). Because *tan* lies on the X chromosome and males are hemizygous, I tested hybrid females (mothered by *D. yakuba*) that either contained a mutant *D. santomea* or *D. yakuba* copy of *tan* (**Figure 2.4E, F**). Offspring bearing a functional *D. yakuba tan* allele showed significantly higher pigmentation levels than those with the *D. santomea* allele

(**Figure 2.4G-K**). These data validate the reciprocal hemizygosity test in the *D. yakuba*/*D. santomea* pair, even when using females, which display less extensive pigmentation than do males.

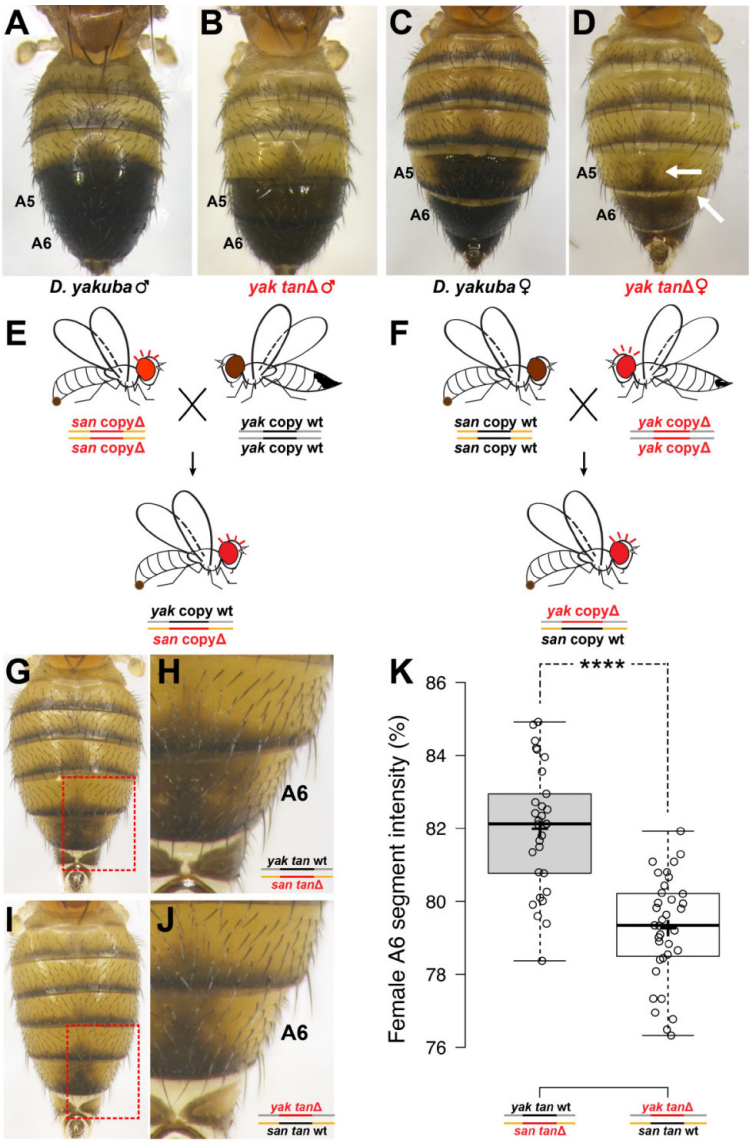

Figure 2.4 Confirmation of the causative role for *tan* in generating pigmentation differences between *D. yakuba* and *D. santomea*. (A-D) Phenotypic comparison between wildtype (A, C) and homozygote *tan* CRISPR flies (B, D) in males (A, B) and females (C, D) of *D. yakuba*. (E, F) Scheme of *D. yakuba*-mothered reciprocal hemizygosity crosses. Bright red coloration indicates 3XP3-RFP fluorescence. (G-J) Phenotypes of female hybrids from the *tan* reciprocal hemizygosity test crosses. (H, J) Closeup views of regions outlined in red in (G, I). (K) Box plot showing quantifications of A6 segment darkness intensity of the female hybrids from the crosses reveal a significant difference between the two groups (Student's T-Test, $p<0.0001$).

The regulatory architecture of the BX-C has been extensively studied in *D. melanogaster* (Maeda and Karch, 2006), comprising an array of CREs between *abd-A* and *Abd-B*, each dedicated to activating expression in a specific body segment (**Figure 2.5A**). I observed differences in Abd-B expression mainly in abdominal segment 5 (A5). The *infra-abdominal-5* (*iab-5*) region contains a CRE responsible for Abd-B expression in the embryonic A5 segment, and mutations within this region result in defective A5 identity (Celniker et al., 1990; Hopmann et al., 1995). Within the ~14 kb that comprise the *iab-5* region, previous research identified a 1kb initiator element, "IAB5", that is sufficient to drive expression in the posterior segments of *D. melanogaster* embryos (Busturia and Bienz, 1993)(**Figure 2.5A**). I therefore evaluated the possibility that IAB5 participates in the divergence of Abd-B expression.

Regulatory alleles removing the approximately 1kb IAB5 initiator element of *iab-5* were replaced in both *D. yakuba* and *D. santomea* with a *3xP3*::*DsRed* (RFP) cassette using CRISPR/Cas9 assisted homology directed repair (Sander and Joung, 2014; Yu et al., 2013) (**Figure 2.5A**). *D. yakuba* individuals heterozygous for the IAB5Δ deletion exhibited reduced pigmentation in the A5 body segment, confirming the conserved function of this region (**Figure 2.6B**) (Mihaly et al., 2006). Individual *D. yakuba* homozygous for the IAB5Δ mutation showed the complete loss of A5 pigmentation, as well as some defects in the A6 segment, suggesting that IAB5 is required in both A5 and A6 body segments (**Figure 2.6C**), which is consistent with the regulatory mechanism of BX-C (**Figure 1.6**) (Lewis, 1978). In the reciprocal hemizygosity test, male hybrids with a functional *D. yakuba* IAB5 allele showed subtle yet significantly higher A5 segment pigmentation than males bearing a functional *D. santomea* IAB5 allele (**Figure 2.5B-H**). To substantiate that the observed pigmentation difference between reciprocal hemizygotes was caused by altered *Abd-B* regulation, I assayed their pupal Abd-B expression. Offspring containing the *D. yakuba* functional allele of IAB5 have sporadic Abd-B expression in A5 epithelial

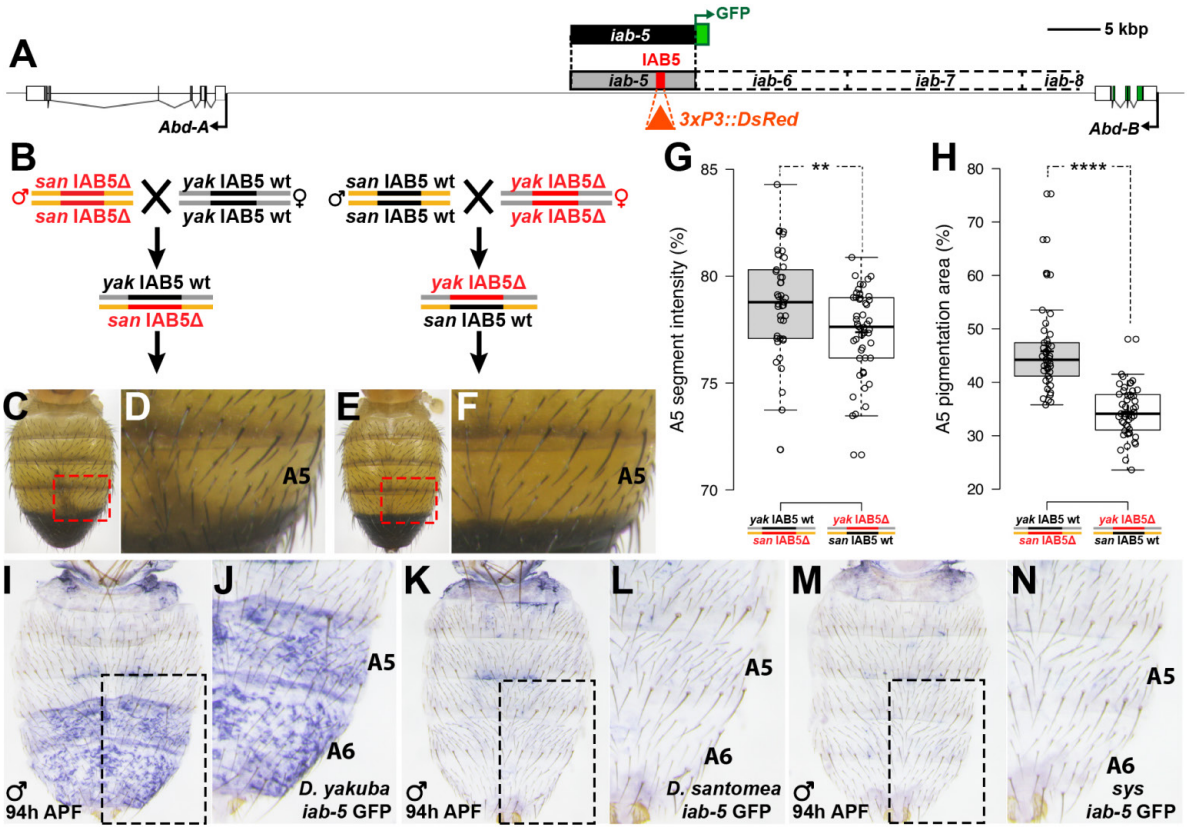

Figure 2.5 Changes to the *cis*-regulatory region of *Abd-B* contribute to its expression and phenotype in *D. santomea*. (A) Map of the *abd-A* and *Abd-B* genomic region, indicating the positions of tandem *infra-abdominal* regions, including the *iab-5* region highlighted in grey. The 1 kb IAB5 initiator element is shown in red. The orange triangle indicates the replacement of IAB5 by a *3xP3*::*DsRed* (RFP) cassette through CRISPR/Cas9 assisted homologous recombination. (B) Scheme for the reciprocal hemizygosity test of the IAB5 deletion mutants. (C-F) Representative abdominal images of the adult progeny from crosses in (B). (C, D) Hybrid animal bearing a wild type *D. yakuba* IAB5 copy. (E-F) Hybrid animal bearing a wild type *D. santomea* IAB5 copy. (G, H) Measurement of A5 segment pigment intensity (G) and A5 pigmented area (H) in male progeny of the crosses shown in (B) (Student's T-Test, ****: $p<0.0001$, **: $p<0.01$). (I-N) *In situ* hybridization to compare GFP mRNA patterns from the *iab-5* GFP reporter transgenic flies at 94 hAPF. (I-L) Abdominal activities of the *iab-5* reporter constructs from *D. yakuba* (I, J) and *D. santomea* (K, L). (M, N) Abdominal activity of the *D. santomea iab-5* reporter construct in which the initiator element is replaced by *D. yakuba* version (*sys iab-5* GFP).

cells, while those bearing the wildtype *D. santomea* IAB5 element showed little to no detectible expression in the A5 segment (**Figure 2.2F-G**). These results confirm that changes in *cis* to *Abd-B* contributed to the *D. santomea* pigmentation and Abd-B expression phenotypes.

My reciprocal hemizygosity assay results suggested that alterations in the 1kb IAB5 CRE region, or perhaps in a region that interacts with IAB5, altered Abd-B expression in *D. santomea*. To localize the causative changes and identify the regulatory sequences necessary for late pupal abdomen expression of Abd-B, I first performed reporter assays in which the minimal 1kb IAB5 initiator element from *D. yakuba* and *D. santomea* were evaluated for their ability to activate a GFP reporter gene's expression (**Figure 2.6D**). To detect subtle differences in activity, I directly compared these regulatory regions in a common genetic background in which the reporter genes were individually inserted into defined sites of the *D. melanogaster* genome. While these constructs drove identical posterior expression patterns in the embryo (**Figure 2.6E, F**), these patterns did not persist into pupal development (**Figure 2.6G, H**).

The IAB5 minimal initiator element is contained within a larger 13.8 kb region of DNA named *iab-5* which has at least one maintenance element required to drive proper Abd-B expression into larval development (Maeda and Karch, 2006). I cloned the entire ~15 kb region orthologous to *iab-5* from *D. yakuba* and *D. santomea* into our reporter system (**Figure 2.5A**). Although these constructs drive identical expression patterns in embryonic stages (**Figure 2.6J, K**), I observed slight differences in the intensity of GFP in the pupal abdomen (**Figure 2.6L, M**). Because GFP protein is known to persist after gene expression has terminated (Arnone et al., 2004), I measured GFP mRNA during pupal development to determine whether these constructs exhibit differences in transcriptional maintenance. Although expression from the *D. yakuba* reporter was detectible in the A5 and A6 body segments at 94 hAPF (**Figure 2.5I, J**), the *D. santomea* construct failed to produce GFP mRNA at this stage (**Figure 2.5K, L**).

Altogether, my results indicate that changes in *cis* to the Hox gene *Abd-B* contribute to the reduced abdominal pigmentation of *D. santomea*. Moreover, at least some of these differences are located in the 15 kilobase *iab-5* region that affect its maintenance function.

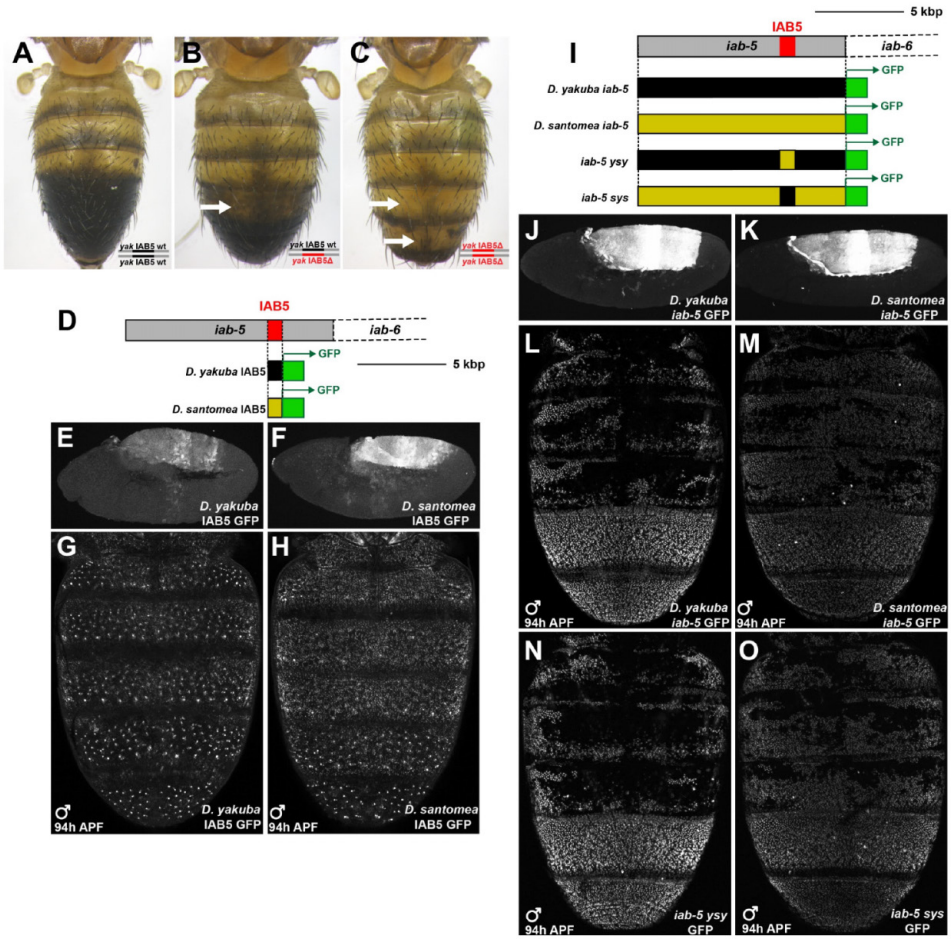

Figure 2.6 Analysis of the abdominal pigment patterning function of the *iab-5* region. (A-C) Phenotypic comparison among wild type (A), IAB5Δ heterozygote (B) and IAB5Δ homozygote male flies (C) in *D. yakuba*. White arrows indicate segments with reduced pigmentation. (D) Schematic of the 15 kb *iab-5* region, showing the 1kb IAB5 initiator element which was cloned from *D. yakuba* and *D. santomea* into reporter constructs. (E-H) IAB5 GFP reporter constructs from *D. yakuba* and *D. santomea*, showing activity in embryos (E, F) and late staged pupal abdomens (G, H). (I) Schematic of the entire 15kb *iab-5* regulatory region GFP reporter constructs from *D. yakuba* and *D. santomea* and the chimeric constructs of the *D. yakuba iab-5* region, with the initiator element IAB5 replaced by *D. santomea* (*iab-5 ysy*), and its reciprocal construct (*iab-5 sys*). (J, K) The activities of *D. yakuba* and *D. santomea* whole *iab-5* regulatory regions in embryos. (L-O) Activity of the four constructs schematized in (I) in the abdomens at 94 hAPF.

2.4 IDENTIFICATION OF Pdm3 AS A CAUSATIVE FACTOR ON THE SECOND CHROMOSOME

The changes I identified in *Abd-B* are expected to have a large effect on *yellow*, but they only partly explain the absence of *yellow* expression in the *D. santomea* abdomen, because the previously-identified QTL on chromosome two also alters *yellow* expression (**Figure 2.1J**). The Introgression II region resides within the chromosome II QTL (**Figure 2.1R**), and includes a gene that has been previously implicated in tergite pigmentation and evolution. In a previous RNAi screen of transcription factors that control abdominal pigmentation, the *pdm3* gene was identified as a pigment suppressing factor (Rogers et al., 2014). *pdm3* encodes a POU domain transcription factor that was initially identified for its roles in the specification of olfactory receptor neurons (Chen et al., 2012; Tichy et al., 2008). Its role in pigment evolution was recently implicated in a correlational study of multiple instances of female-limited abdominal color dimorphism in the montium clade (Yassin et al., 2016).

I compared the distribution of Pdm3 expression in the pupal abdomen of *D. yakuba* to *D. santomea* (**Figure 2.7B, C**). In *D. yakuba*, Pdm3 is broadly expressed during early stages of pupal development, and persists within the pleura, which form flexible membranes between the stiff abdominal tergites (**Figures 2.7B and 2.8C**). Notably, expression of Pdm3 subsides in the *D. yakuba* abdomen during late development (**Figure 2.8E, G and I**), when many pigmentation enzymes are activated. In *D. santomea*, although the early patterns of Pdm3 expression mirror those of *D. yakuba* in the third instar brain (**Figure 2.8A, B**), expression in the epithelium underlying posterior tergites, especially A6 is consistently higher (**Figure 2.7C**), and both A5 and A6 expression persist through late pupal development (**Figure 2.8D, F, H and J**).

To determine whether evolution of *pdm3* has contributed to the *D. santomea* phenotype, CRISPR/Cas9-mediated homology-directed repair was used to generate *pdm3* null mutations marked with a a *3xP3*::*DsRed* (RFP) in *D. santomea* and *D. yakuba* (**Figure 2.7A**) for reciprocal

hemizygosity testing. While male progeny of the reciprocal hemizygous crosses did not differ in phenotype when mothered by *D. yakuba* (**Figure 2.8K-M**), I found significant differences in females (**Figure 2.8N-P**). The lack of phenotypic differences in males may be due to epistatic effects of the hybrid background which has a strong pigmentation phenotype and could therefore mask subtle effects of *pdm3*. To test the potential effects of *pdm3* in a different genetic background, I performed reciprocal hemizygosity tests for *pdm3* in *D. santomea*-mothered hybrids, which have a lighter pigmentation phenotype due to the presence of the *D. santomea* X chromosome (**Figure 2.7I, J**). These reciprocal hemizygotes displayed large pigmentation differences (**Figure 2.7D-H**), implicating genetic changes at the *D. santomea pdm3* gene that enhanced its pigment-suppressing role in this species.

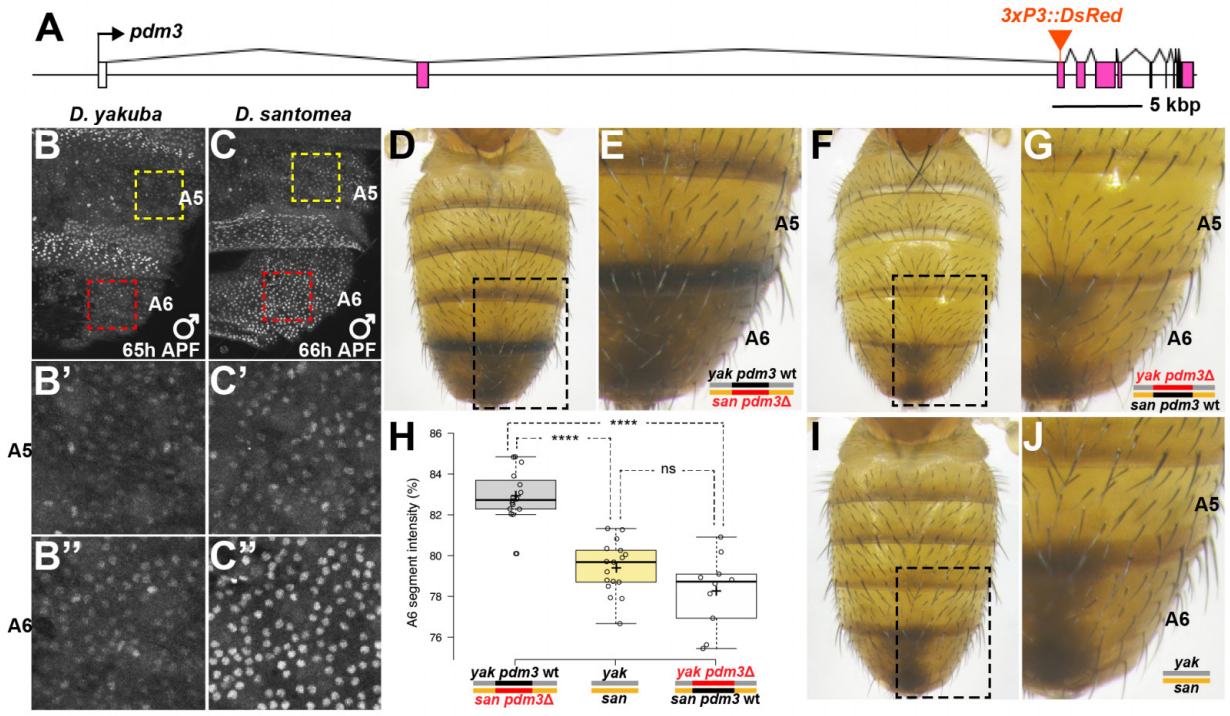

Figure 2.7 Identification of *pdm3* as a second chromosome locus that contributes to the *D. santomea* phenotype. (A) Schematic of the *pdm3* locus. The orange triangle indicates the position of a *3xP3*::*DsRed* cassette inserted to generate a *pdm3* null mutation. (B, C) The comparison of Pdm3 expression between *D. yakuba* and *D. santomea* in male pupal abdomens at 65-66 hAPF. (B', B", C' and C") show zoomed-in areas outlined in (B, C). (D-H) Phenotypes of male hybrids from the *D. santomea*-mothered *pdm3* reciprocal hemizygosity test crosses. (E, G) show zoomed-in regions from black dashed-line boxes in (D, F). (H) Boxplot showing the quantifications of A6 segment intensities among the *D. santomea*-mothered *pdm3* RHT crosses and *D. yakuba*/*D. santomea* hybrids (Student's T-Test, ****: $p<0.0001$, ns: not significant). (I, J) shows the phenotype of a *D santomea*-mothered hybrid male.

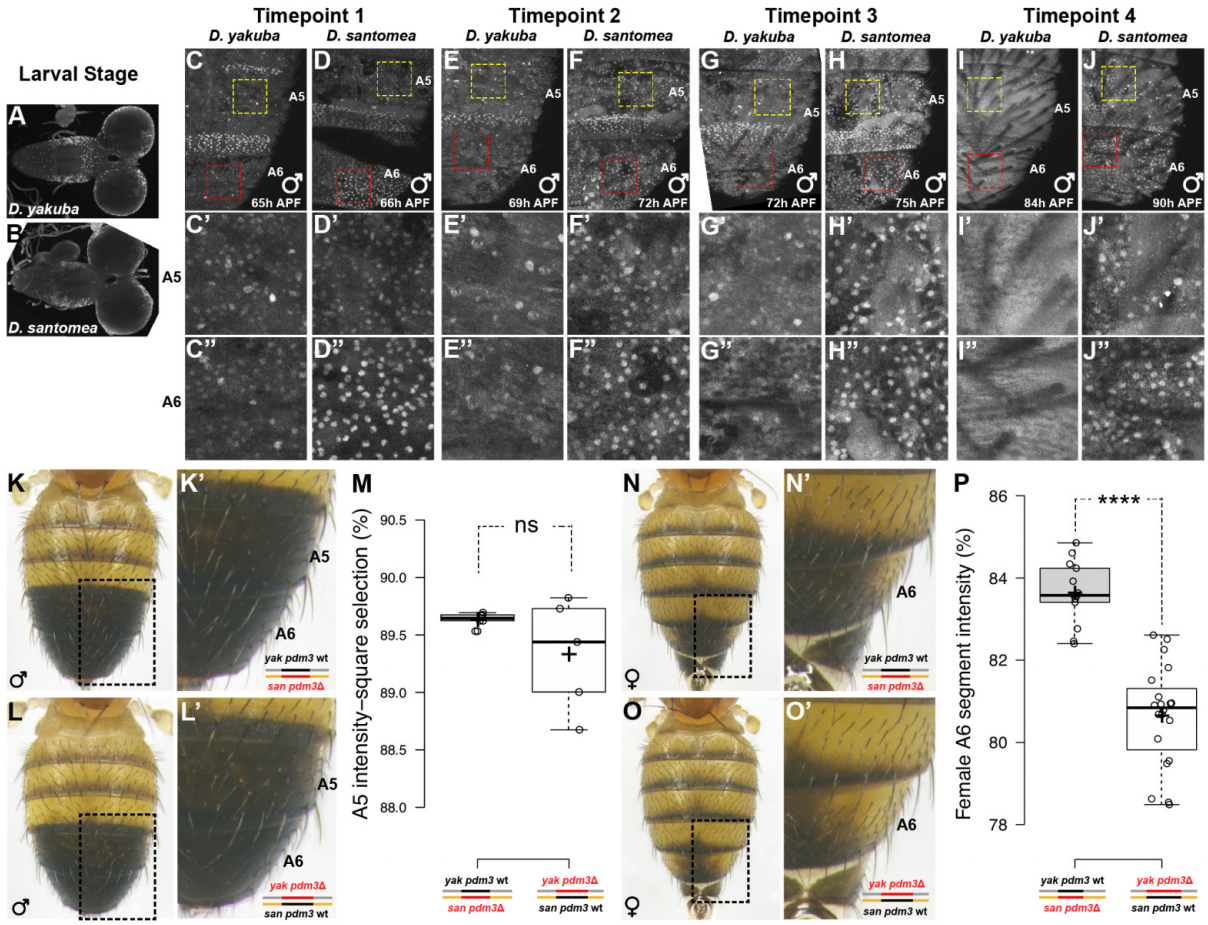

Figure 2.8 Timecourse of Pdm3 expression differences and its phenotypic consequences by reciprocal hemizygosity testing. (A-J) Time course of Pdm3 expression, showing the third instar larval brain, and different time points of pupal development of the *D. yakuba* and *D. santomea* abdomen. (C'-J', C"-J") show zoomed-in images of A5 and A6 body segment expression presented in (C-J). (K-L) Phenotypes of male hybrids from the *pdm3* RHT crosses. (K', L') show zoomed-in regions from dashed-line boxed in (K, L). (M) Quantification of A5 segment pigment intensity of the male reciprocal hemizygotes show no significant differences (Student's T-Test). (N-O) Phenotypes of female hybrids from the *pdm3* RHT crosses. (N', O') show zoomed-in regions from the outlined boxes in (N, O). (P) Quantifications of A6 segment pigment intensity of the female hybrids show significant differences between the two groups of crossings (Student's T-Test. $p<0.0001$).

2.5 A TRANSPOSON INSERTION AT *ebony* INCREASED EXPRESSION IN *D. SANTOMEA*

Introgression IIIB contains a homozygous *D. yakuba* region (**Figure 2.1Q**) that includes the pigment-reducing gene *ebony*. The expression of *ebony* is altered in this strain (**Figure 2.1N**), suggesting that the *cis* regulatory region of *ebony* itself has evolved. To test whether *ebony* contributed to the pigmentation difference between *D. yakuba* and *D. santomea*, I performed a reciprocal hemizygosity assay using *null ebony* alleles of both species (**Figure 2.9A-D**). Similar to my results with *pdm3*, *ebony* reciprocal hemizygotes mothered by *D. yakuba* females showed a fully pigmented A5-A6 dark phenotype that was not significantly different (**Figure 2.9E-I**). I also obtained a small number of progeny from a *D. santomea* mothered cross. In this case, the hybrids containing the *D. yakuba* functional allele displayed darker A5 body segments compared to those with *D. santomea* alleles (**Figure 2.9J-N**). To better observe the phenotypic difference in the posterior abdomen, we generated an *ebony* null allele in the Introgression IIIB line to perform the reciprocal hemizygosity test in a sensitized background in which the only *D. yakuba* genes derive from the *ebony* region. I crossed wildtype and mutant introgression lines to mutant and wildtype *D. santomea* strains, respectively. The progeny from the cross of Introgression IIIB to *D. santomea ebony* mutants contain a *D. yakuba* functional *ebony* allele. These progeny produced darker A5 and A6 body segment intensities compared to the flies with the *D. santomea* functional allele (**Figure 2.10F-K**), confirming the phenotypic contribution of changes at *ebony* in *D. santomea.*

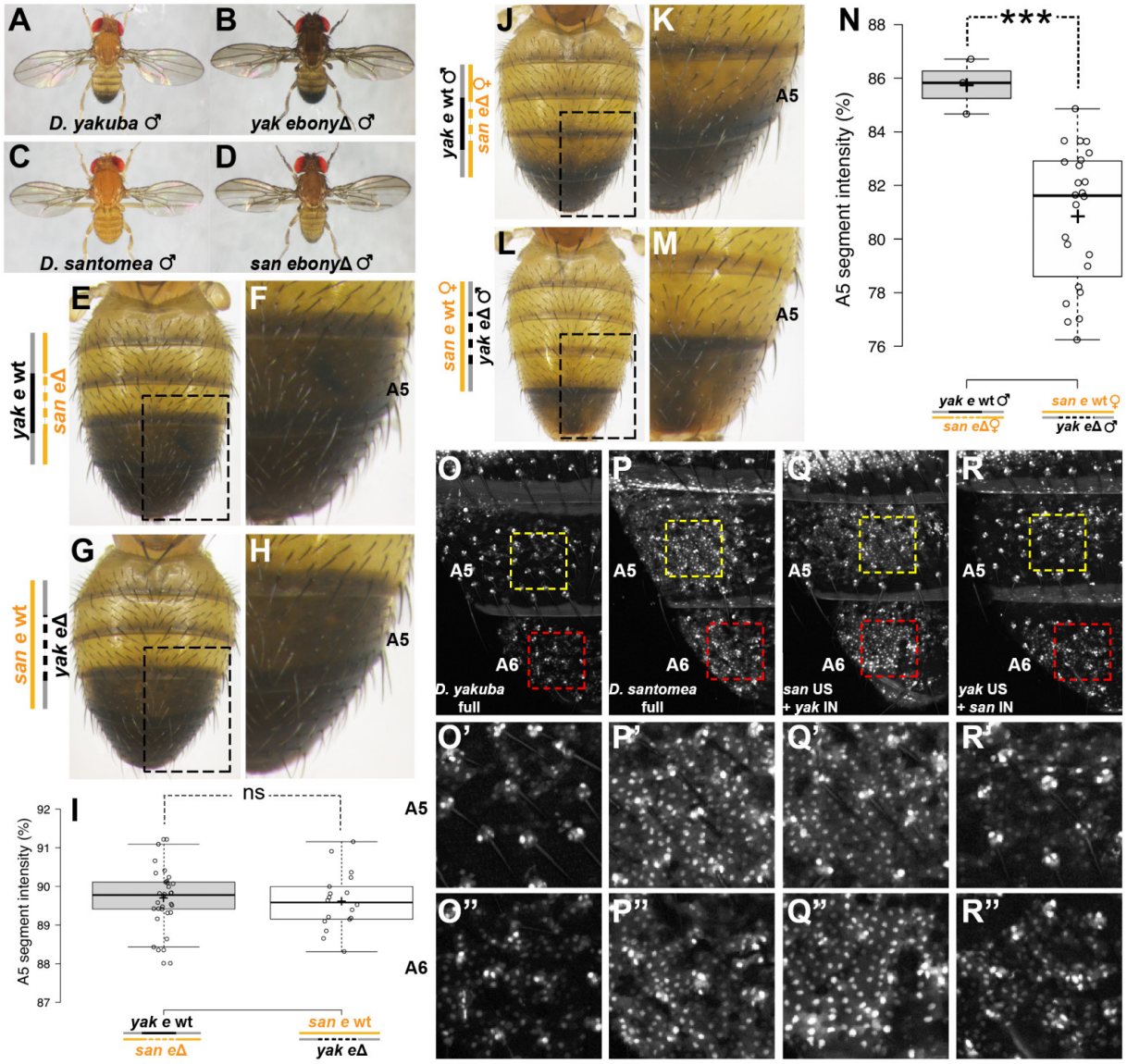

Figure 2.9 Measuring the phenotypic impact of variation at *ebony*, and localizing the source of its regulatory changes. (A-D) Phenotypic comparison between wild type (A, C) and *ebony* CRISPR homozygote male flies (B, D) in *D. yakuba* and *D. santomea*. (E-H) Phenotypes of male hybrids from the *D. yakuba*-mothered *ebony* reciprocal hemizygosity test crosses. (F, H) show zoomed-in regions from black outlined regions in (E, G). (I) Box plot of A5 segment intensities of *D. yakuba*-mothered male progeny from *ebony* RHTs doesn't show a significant difference (Student's T-Test). (J-M) Phenotypes of male hybrids from the *D. santomea*-mothered *ebony* RHT crosses. (K, M) show zoomed-in regions from outlined regions in (J, L). (N) Box plot of A5 segment pigment intensity from rare *D. santomea*-mothered *ebony* reciprocal

hemizygotes reveals a striking difference in pigmentation (Student's T-Test, $p < 0.001$). (O-R) Reporter activities of entire regulatory region of *ebony* from *D. yakuba* (O), *D. santomea* (P) and the intron-swapped chimeric constructs (Q, R) from the two species. (O'-R', O"-R") show zoomed-in regions from yellow and red outlined regions in (O-R), indicating A5 (O'-R') and A6 (O"-R") segment activities.

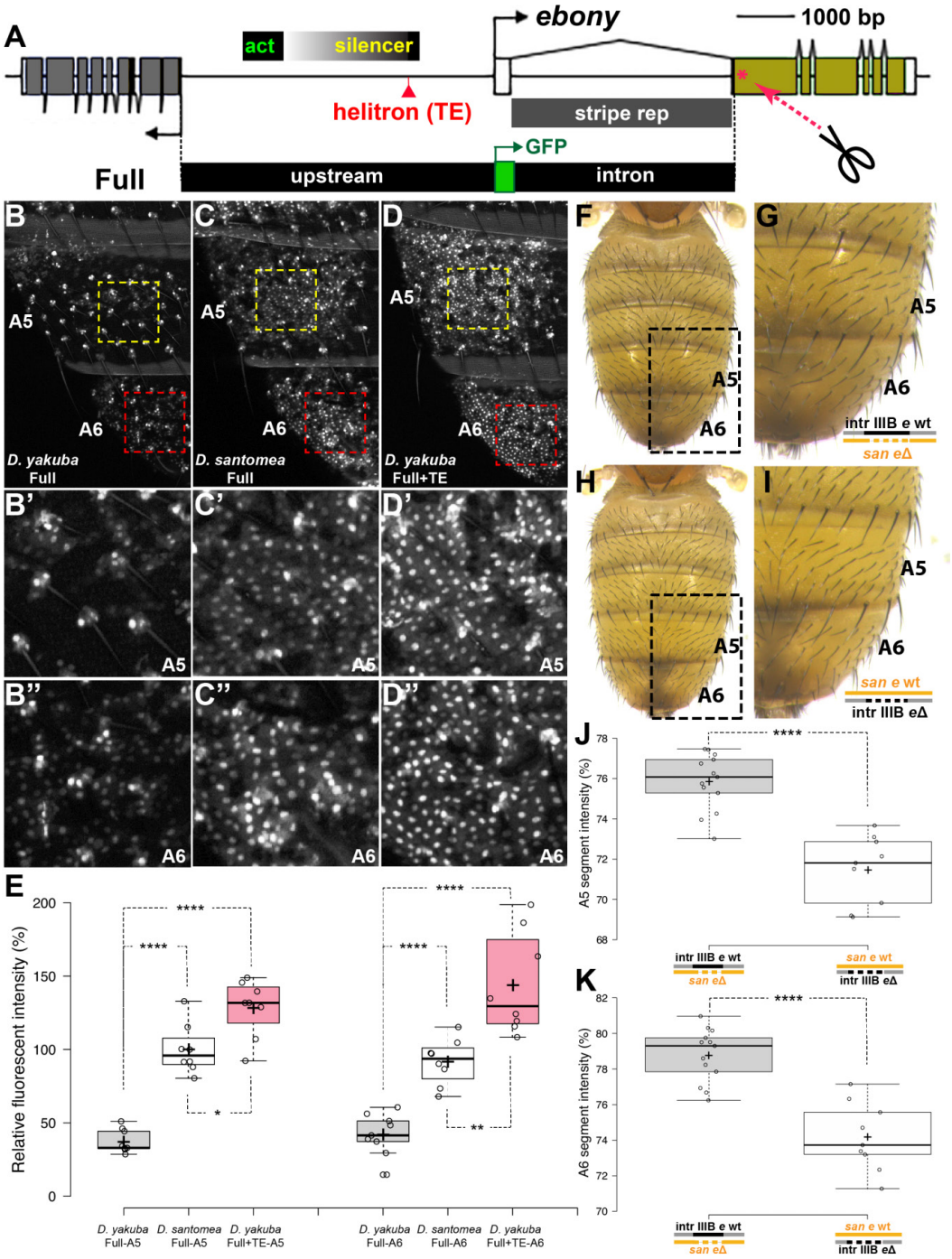

Figure 2.10 A transposon insertion at *ebony* conferred increased expression in the *D. santomea* abdomen. (A) Map of the *ebony* locus with positive-acting abdominal enhancer ("act") and two repressing regions, "silencer" which restricts expression from male A5 and A6 segments, and the intronic "stripe rep" element which prevents expression at the posterior edges of anterior body segments. Schematic of GFP

reporter Constructs “Full” that include the upstream and intronic regions are depicted. Compared to the *D. yakuba* upstream regulatory region, *D. santomea* has a ~500 bp transposon insertion (red triangle). Red star indicates the location of guide RNAs used to generate *ebony* mutants through non-homologous end joining. (B-D) GFP expression patterns of transgenic reporters in the posterior body segments A5 and A6. (B) *D. yakuba* “*ebony* Full” construct. (C) *D. santomea* “*ebony* Full” construct. (D) “*yak ebony* + TE” construct, in which the ~500 bp helitron element from *D. santomea* was precisely cloned into the *D. yakuba* silencer element. (B’-D’, B”-D”) Closeup images outlined in (B-D) by yellow and red dashed lines. (E) Boxplot of quantifications of relative fluorescent intensities of A5 and A6 abdominal segments among the three reporter constructs (Student’s T-Test, ****: $p<0.0001$, **: $p<0.01$, *: $p<0.05$). (F-K) Reciprocal hemizygosity test for *ebony* performed in a *D.santomea/*Introgression IIIB background. (F, G) Representative male abdomen of the progeny from the cross of Introgrssion IIIB to *D. santomea ebony*Δ containing a *D. yakuba* functional ebony allele. (H, I) Representative male abdomen of the progeny from the cross of Introgression IIIB *ebony*Δ to *D. santomea* wild type, which has a functional allele of the *D. santomea ebony* gene. (G, I) show zoomed-in views of the regions outlined in (F, H). Boxplots of A5 (J) and A6 (K) segment intensities revealed significant differences between the two crosses (Student’s T-Test, ****: $p<0.0001$).

Considering how *ebony* expression in Introgression IIIB is more similar to *D. yakuba* rather than *D. santomea*, correlating with its pigmentation phenotype, I speculated that the *D. santomea ebony* gene has evolved through changes to its *cis*-regulatory region. Previous work on *ebony* revealed multiple CREs controlling *ebony* transcription in the *D. melanogaster* abdomen (Rebeiz et al., 2009a). An abdominal-expression activating enhancer, which is located 2.5 kb 5' of the promoter ("act." in **Figure 2.10A**), promotes *ebony* expression throughout the abdomen. Two silencers spatially restrict the broad activity of the enhancer in the abdomen (**Figure 2.10A**). A region 5' of the promoter, the male repression element ("silencer" in Figure 2.10A), is required to restrict expression from pigmented A5 and A6 segments. A second region, located in the first intron, represses activity in the epidermis underlying the posterior fringes of each tergite. The expanded expression phenotype of *D. santomea ebony* corresponds with the regions where these silencers function.

I therefore assessed the function of the *ebony cis*-regulatory sequences in reporter assays that included all relevant enhancer and silencer elements (**Figure 2.10A**). The *D. yakuba ebony* construct shows weak activity in the A5 and A6 segment epithelial cells (**Figure 2.10B**), recapitulating its *ebony* expression pattern, while the *D. santomea* construct exhibits increased activity (**Figure 2.10C**), mimicking *D. santomea*'s *ebony* pattern. Flies bearing the *ebony* 5' region from *D. santomea* fused to the *D. yakuba* intron show a posterior expansion of GFP (**Figure 2.9Q**), indicating that the expanded activity stems from alterations to a 5' regulatory element in *D. santomea*.

Sequencing of the *D. yakuba* and *D. santomea ebony* 5' regions uncovered a ~500 bp transposable element of the helitron class (Kapitonov and Jurka, 2007) inserted in the presumptive location of the male repression element of *D. santomea* (**Figure 2.10A**). We were curious if this insertion was fixed in *D. santomea*. A survey of 128 iso-female lines of *D. santomea* showed that all lines contained the same-sized insertion by a PCR test. Conversely, of the 102 *D. yakuba* strains tested, only two had an insertion related to the *D. santomea* helitron element. From

these data, the 95% confidence intervals for the frequency of the transposon insertion is 0 to .036 for *D. yakuba*, and .97 to 1 for *D. santomea* (exact binomial test). To test the ramifications of this fixed insertion, I generated a *D. yakuba ebony* regulatory region reporter containing the *D. santomea* 500bp insertion. This modified construct had increased A5 and A6 segmental activity (**Figure 2.10D, E**), confirming the significance of this insertion in generating the *D. santomea ebony* expression phenotype.

2.6 *CIS*-REGULATORY EVOLUTION AT *yellow* ATTENUATED ITS ABDOMINAL CRE

Although we recovered no introgressions that contained the *yellow* gene, the previous QTL analysis by Carbone et al. had identified a small QTL peak near the telomere of the X chromosome that corresponds to the position of this critical pigmentation gene (Carbone et al., 2005). Previous work had shown that the *D. santomea* body element CRE of *yellow* has maintained its ability to activate reporter gene expression in the abdomen (Jeong et al., 2006). This finding was further substantiated by the observation that *D. yakuba*/*D. santomea* hybrid males, which contain the *D. santomea* X chromosome display detectible *yellow* expression in the A5 and A6 segments (Jeong et al., 2008). However, the previous reporter study was performed using P-element transgenesis that placed transgenes randomly throughout the genome, making subtle quantitative differences in activity difficult to detect. Thus, I explored the possibility that subtle genetic changes at *yellow* may underlie the minor effect on the X chromosome.

Null mutations of *yellow* in *D. santomea* and *D. yakuba* were generated in order to perform reciprocal hemizygote testing. As with *tan*, the presence of this gene on the X chromosome limited the use of this test to females, which showed a quantitative shift consistent with a subtle effect of this gene on pigmentation (**Figure 2.11B-E, Q**). Examination of *yellow* expression in reciprocal hemizygote pupae revealed differences that correlate with the difference in pigmentation (**Figure 2.11F-I**), suggesting that the phenotype is caused by *cis*-regulatory mutations at *yellow*. To

explore whether this effect could be caused by changes in the characterized abdominal enhancer of *yellow* (Wittkopp et al., 2002b), I tested regulatory elements from *D. yakuba* and *D. santomea* in GFP reporter constructs that were inserted into an identical genomic location. Previous work identified an abdominal "body" element that is active throughout the abdomen (**Figure 2.11A**), and is upregulated in posterior body segments of males (Jeong et al., 2006; Wittkopp et al., 2002b). Next to this element is a wing enhancer that also contains sequences that drive expression in stripes at the posterior edge of each segment (**Figure 2.11A**) (Camino et al., 2015; Roeske et al., 2018; Wittkopp et al., 2002b). I first tested a reporter construct that contains both wing and body enhancer elements. Consistent with a minor effect of *yellow* on the *D. santomea* phenotype, I observed that although the *D. santomea wing-body* reporter drove expression in the posterior abdomen of males and females, it was reduced compared to the orthologous segment tested from *D. yakuba* (**Figures 2.11J-P, 2.12B-D**). To test whether this difference could be localized to the body element, I examined the activity of body element reporters from both species, which showed similar activity differences to the *wing-body* element (**Figure 2.12**). These results suggest that variation within the body element enhancer attenuated its activity, explaining the minor QTL effect observed in this region.

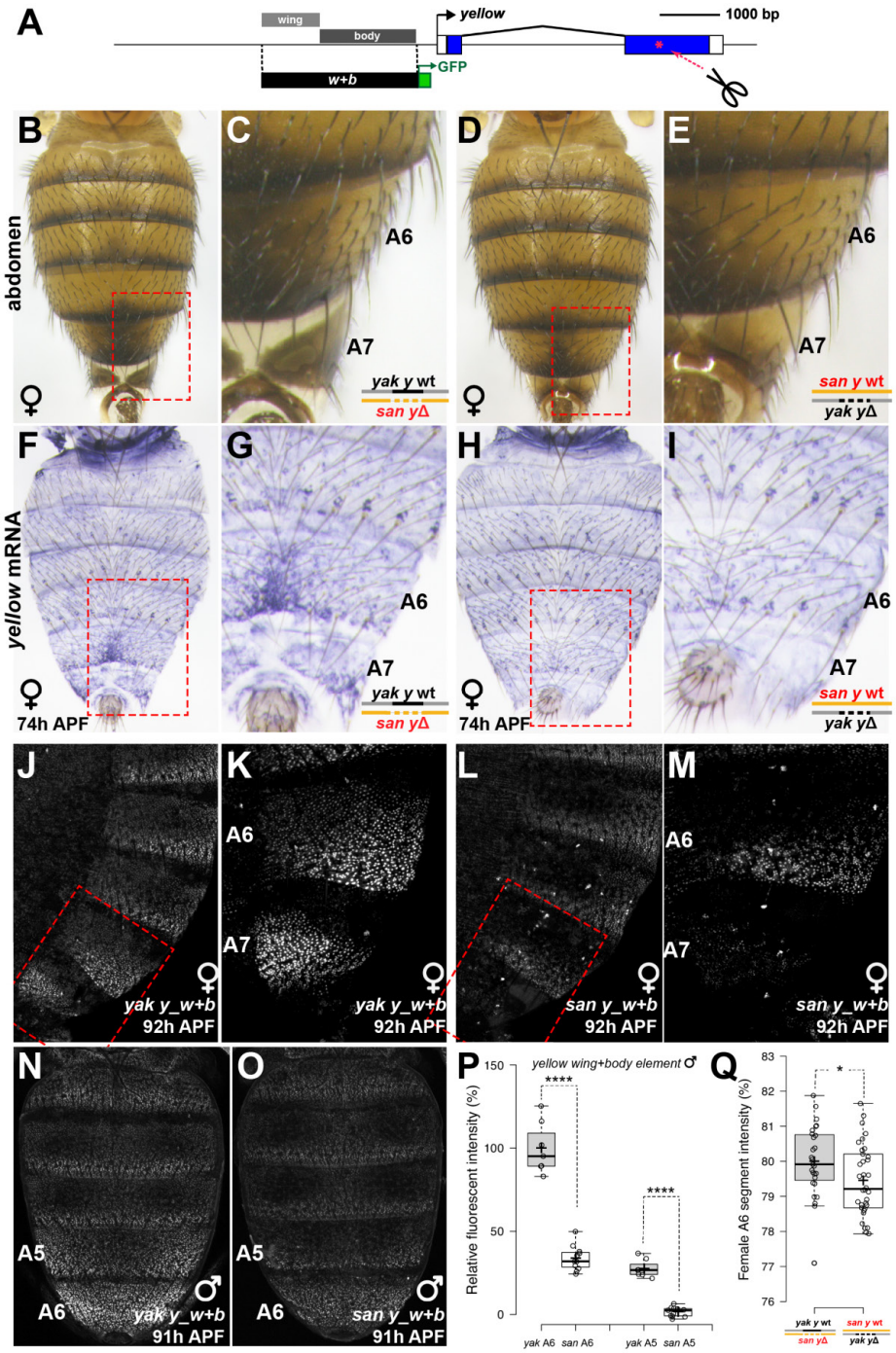

Figure 2.11 *cis* regulatory evolution at *yellow* attenuated its abdominal enhancer in *D. santomea*. (A) Schematic of the *yellow* locus, showing the positions of the wing and body enhancer elements upstream of the transcription start site and the wing+body element GFP reporter construct. Red star indicates location of a guide RNA used to generate the CRISPR mutants through non-homologous end joining. (B-E)

Phenotypes of the female hybrids from the *yellow* reciprocal hemizygosity test crosses. (C, E) Closeup views of the regions outlined in (B, D). (F-I) *In situ* hybridization of female samples from the reciprocal hemizygosity test crosses, showing the accumulation of *yellow* mRNA at 74 hAPF. The zoomed-in images (G, I) are the areas outlined by red dashed lines in (F, H). (J-M) Comparison of *D. yakuba* (J, K) and *D. santomea* (L, M) *yellow* wing+body reporter activities in the female A7 segment at 92 hAPF. (K, M) Closeup images of the outlined regions in (J, L). (N-P) Examination of *yellow* wing+body element activities in males. Boxplot in (P) shows significant differences both in A5 and A6 segments between *D. yakuba* and *D. santomea* constructs (Student's T-Test, ****: $p<0.0001$). (Q) Quantifications of female hybrid A6 body segment intensities between the two crosses from *yellow* reciprocal hemizygosity test revealed a slight difference (Student's T-Test, *: $p<0.05$).

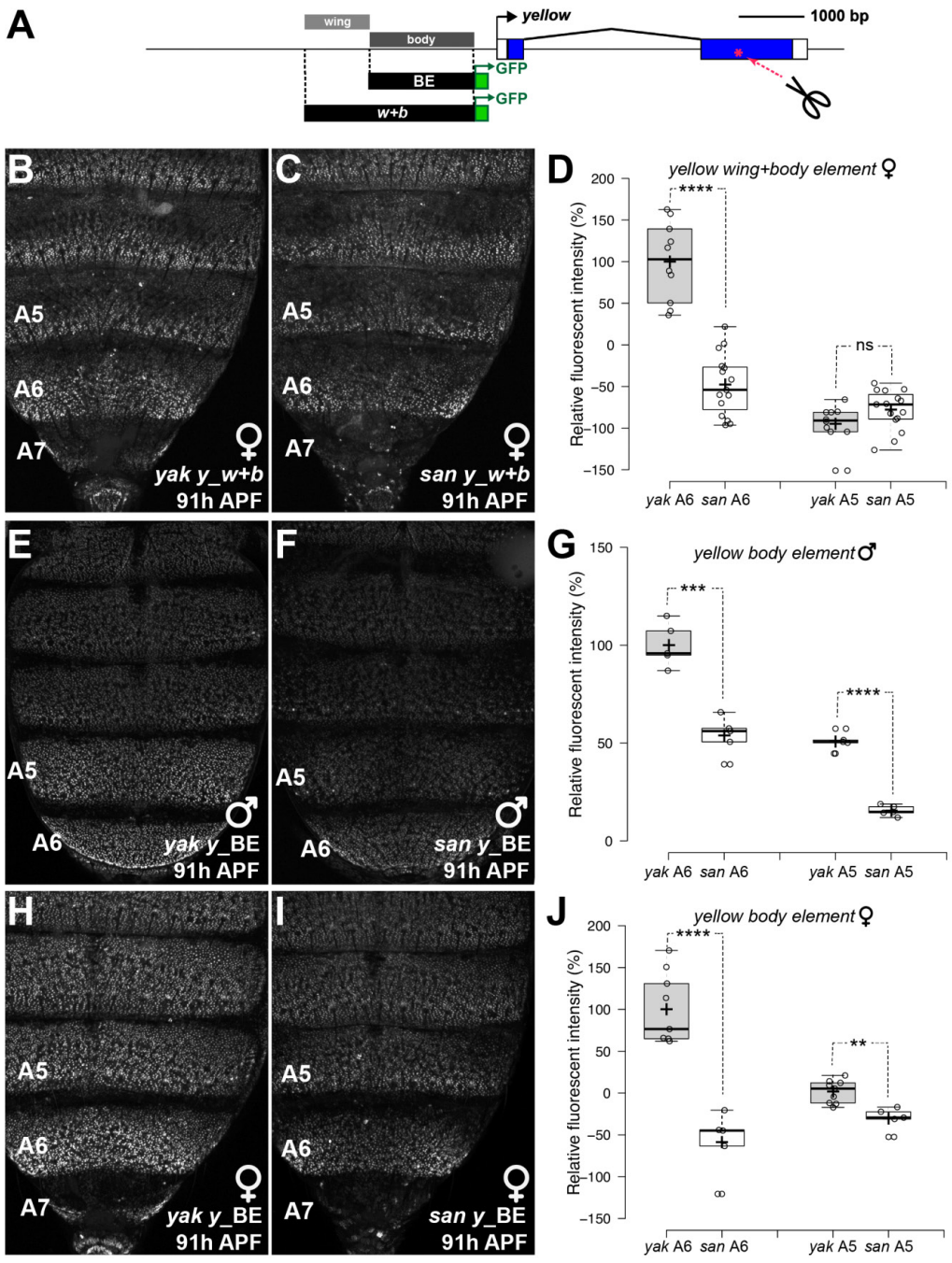

Figure 2.12 Localization of activity-causing differences in the *yellow* regulatory region to the abdominal body element. (A) Schematic of *yellow* genomic locus, indicating positions of the wing+body element and body element GFP reporter constructs. (B-D) Comparison of *yellow* wing+body element activities in females. Box plot in (D) reveals a significant difference in reporter intensity among A6 body segments. (E-G) Comparison of *yellow* body element activities in males. Box plot in (G) reveals significant

differences in both the A5 and A6 body segments. (H-J) Examination of *yellow* body element activities in females. Box plot in (J) shows significant differences in both A5 and A6 body segments. Statistical analyses were performed with a Student's T-Test. The number of stars indicates different *p* values. ****: $p<0.0001$, ***: $p<0.001$, **: $p<0.01$, *: $p<0.05$, ns: not significant.

3.0 DISCUSSION, CONCLUSION AND FUTURE DIRECTIONS

In my work, I have dissected a polygenic trait to five causative loci throughout a gene regulatory network, and have demonstrated their contributions to a phenotype that constitutes a difference in body plan. These loci span the different levels of a network's hierarchy, including the Hox gene *Abd-B* that specifies posterior segmental identity from very early stages of development, another transcription factor, *pdm3*, that collaborates with *Abd-B* to modulate the output of the network, and three terminal genes whose encoded enzymes directly catalyze catecholamine-based pigment formation. Of note, all the identified loci exhibit spatio-temporal expression differences between the two sister species, and in four of the genes, I identified causative alterations in their CREs. My work illustrated a near-complete dissection of a polygenic trait that serves as an important model for how body plans evolve over longer macroevolutionary timescales.

3.1 THE NEAR COMPLETE DISSECTION OF A POLYGENIC TRAIT

One of the major challenges with traditional quantitative genetic approaches is the identification of the causative genes underlying QTL peaks, due to the low resolution of the large intervals in QTL maps (Georges, 2007; Mackay et al., 2009). Based on deep sequencing and alignment, the introgression mapping approach I employed narrowed down the phenotype-conferring region to several megabases, from which I could select candidate genes. Matching the phenotypes of introgression lines to the expression of candidate genes within and outside of the introgression intervals allowed me to identify *cis* and *trans* changes in the pigmentation network. Of note, these

expression changes correlated very well with the overall phenotype of each introgression line, suggesting that I can account for the vast majority of causative changes within each introgression interval tested.

For example, the *ebony* expression pattern in introgression IIIB is finely anti-correlated with the brown pigmentation pattern in A6 of this line, greatly supporting the role of the changes I detected in *ebony* for causing this phenotype. Introgression IIIA fully restores A5 expression of Abd-B to the *D. santomea* abdomen, which can account for the minor effects this line has on midline pigmentation. These two lines revealed the separable contributions of *ebony* and *Abd-B* to the QTL peak on Chromosome III. Previous work on *tan* also confirmed its contribution by testing its expression in an X Chromosome introgression line, which closely mirrored the phenotype of this introgression (Rebeiz et al., 2009b). Within the Chromosome II introgression, I identified the transcription factor *pdm3*, whose expression in *D. santomea* correlates extremely well with the phenotype of this introgression line. However, *pdm3* is located near the boundary of the second chromosome QTL peak, leaving open the possibility that additional genes on the second chromosome may also contribute to that QTL. Finally, I have also confirmed *yellow*'s contribution to a small QTL effect on the X Chromosome, based on transgenic reporter assays and the reciprocal hemizygosity test. Given the small effect of this region on the phenotype in the previous QTL study (Carbone et al., 2005), it is unlikely, though possible, that additional genes contribute to this peak. In summary, I have identified at least one causative gene for each of the previously detected QTL, and for two of the four QTL (on Chromosome X and III), the causative genes I found can account for the majority of their respective effects.

3.2 SHIFTS IN NETWORK ARCHITECTURE SHAPE EPISTATIC INTERACTIONS

A significant barrier to the identification and interpretation of phenotype-altering variants is posed by epistatic interactions that mask phenotypic effects in a background dependent manner

(Chandler et al., 2013). My identification of nearly all loci underlying the major effect QTL in *D. santomea* allowed me to examine the molecular logic by which the epistatic interactions among participating loci have manifested. I posited that important interactions exist between *Abd-B* and the other loci. While the elimination of A5 segment expression of Abd-B led to a drastic phenotype in the *D. yakuba* IAB5 deletion line (**Figure 2.6B**), the restoration of its expression to A5 in *D. santomea* (Introgression IIIA) had only a subtle phenotypic effect (**Figure 2.1C**). I hypothesized that this epistatic relationship was due to the multiple connections between *Abd-B* and target genes within the network of *D. yakuba* that have been modified or lost in *D. santomea*. To confirm that the pigmentation network responds to *Abd-B* in a concerted manner in *D. yakuba*, I examined gene expression in heterozygous IAB5 deletion mutants, which display a variably penetrant reduction in A5 pigmentation. In this line, *yellow* and *tan* expression is nearly eliminated from A5, and *ebony* expression expands into the A5 segment (**Figure 3.1**). These multiple changes in the network of *D. yakuba* contrast with the moderated effects of *Abd-B* restoration in *D. santomea*: while some *yellow* expression is restored to the abdomen, the regulatory elements of *ebony* and *tan* that respond to *Abd-B* have been disrupted in the *D. santomea* background. Thus, changes throughout this Hox-regulated network are epistatic in *D. santomea* to the alterations in *Abd-B* that would otherwise generate more pronounced phenotypic effects.

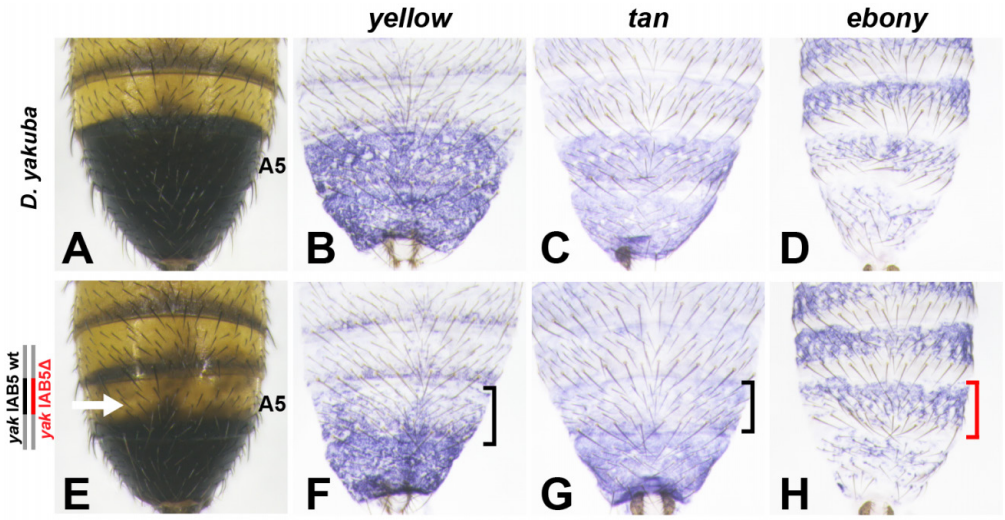

Figure 3.1 Changes to Abd-B expression alter many genes of the pigmentation gene network. Compared to wild type *D. yakuba* (A-D), animals heterozygous for a 1kb deletion of the IAB5 region show reduced A5 pigmentation (E). In these animals, the expression of both *yellow* (F) and *tan* (G) are greatly reduced in the A5 segment. The expression of *ebony* is increased in the A5 segment of heterozygous IAB5 deletion animals (H) compared to wildtype *D. yakuba* (D). As with the A5 phenotype of these heterozygotes, the expression of pigmentation genes is sporadic in the A5 segment.

Multiple lines of evidence further support complex interactions between *Abd-B* and *yellow*, *tan*, and *ebony*. Introduction of the *D. yakuba Abd-B* gene into *D. santomea* restores yellow expression to the A6 body segment, and to a small extent in the A5 segment (**Figure 2.1H**). My data also suggest that the *D. yakuba ebony* gene is ectopically expressed in the A5 segment when placed in the *D. santomea* background in which *Abd-B* is altered (**Figure 2.1N**), consistent with a role for *Abd-B* in repressing posterior body segment expression of *ebony*. Similarly, introgressions containing the *D. yakuba tan* gene restore its expression only to the A6 segment, consistent with the restricted expression of Abd-B to A6 in this background (Rebeiz et al., 2009b). Of note, I have not observed differences in Pdm3 expression when tested in the *D. yakuba Abd-B* containing Introgression IIIA line (**Figure 3.2 A, B**), nor in the *D. yakuba* IAB5 homozygote

deletion line (**Figure 3.2 C, D**). Thus, changes to *Abd-B* led to multiple downstream differences, including *yellow*, *tan* and *ebony*, in the network, differences that are masked by other changes in the *D. santomea* background.

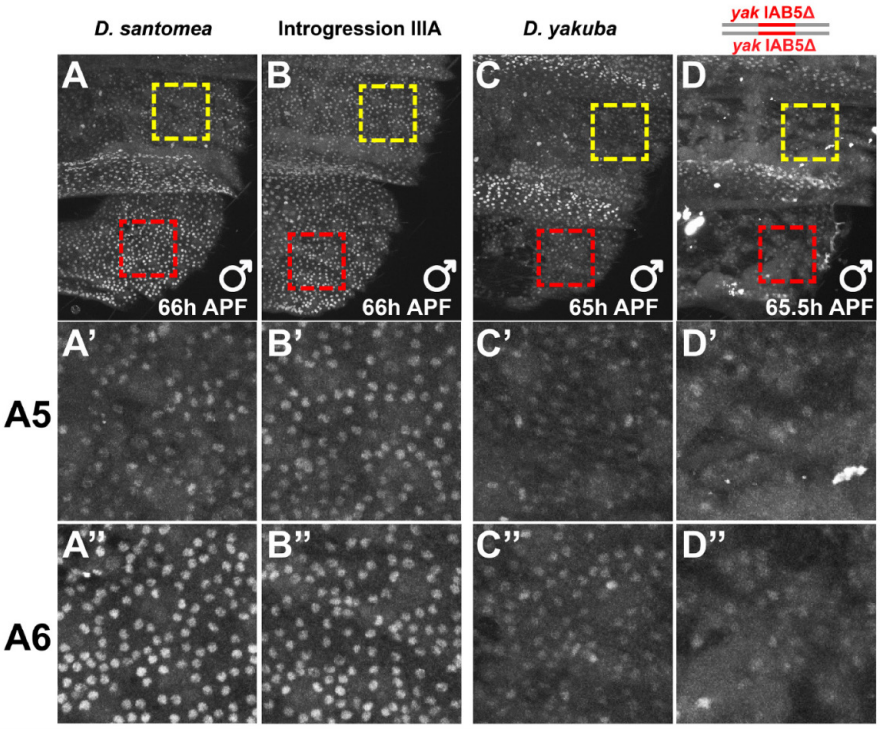

Figure 3.2 No epistatic interactions between *Abd-B* and *pdm3*. *D. santomea* (A) and Introgression IIIA (B) demonstrate similar Pdm3 expression patterns in male pupal abdomens at 66 hAPF. (C, D) Compared to *D. yakuba* (C), animals homozygous for the IAB5 deletion in *D. yakuba* have identical Pdm3 expression patterns in male pupal abdomens at 65 hAPF. (A'-D', A"-D") show zoomed-in areas outlined in yellow (A5) and red (A6) from (A-D).

3.3 IMPLICATING GENES AND REGULATORY ELEMENTS WITH THE RECIPROCAL HEMIZYGOSITY TEST

Advances in genome engineering with the CRISPR/Cas9 system will open many future opportunities to examine the direct phenotypic consequences of genetic variation through tests such as the reciprocal hemizygosity test (Stern, 2014). Using these tools, my data provide a useful precedent for how to analyze *cis*-regulatory divergence in a complex multigenic locus such as the Bithorax complex. Deletion of the IAB5 initiator element was sufficient to change both the phenotype of Abd-B expression (**Figure 2.2F, G**), and pigmentation in reciprocal hemizygous test offspring (**Figure 2.5B-H**). I interpreted this result to indicate the existence of phenotypically relevant variation in this 1kb region, or in a sequence that requires this 1kb region in *cis* for its function. This was supported by reporter assays comparing the 15kb *iab-5* regulatory regions of *D. yakuba* and *D. santomea*. However, a chimeric reporter construct in which the 1kb IAB5 region of *D. santomea* was replaced by the *D. yakuba* sequence (**Figure 2.6I**) reported expression that matched *D. santomea* (**Figure 2.5M, N**). Likewise, a *D. yakuba* construct in which the IAB5 region was replaced by *D. santomea* sequence shows a *D. yakuba* pattern (**Figure 2.6L, N**). Thus, the changes to this region lie outside of the 1kb IAB5 region, but must represent sequences that depend upon IAB5 to exert differential effects on Abd-B expression in *cis*.

While powerful, the reciprocal hemizygosity test is not without caveats, which must be considered, especially when negative results are obtained. For example, my results confirmed a contribution of *tan* to the *D. santomea* phenotype in females, an effect that was not detected when *D. melanogaster tan* mutants were hybridized to *D. santomea* (Matute et al., 2009). Tests of *ebony* and *pdm3* in this assay also failed to confer a detectible A5/A6 phenotype in male hybrids mothered by *D. yakuba* (**Figure 2.8, 2.9**). The most parsimonious interpretation is that the phenotypes were epistatically masked in the *D. yakuba*-mothered hybrid background, as this assay showed anticipated effects for both genes in *D. santomea*-mothered hybrids (**Figures 2.7H,**

2.9N), and in tests involving mutant introgression lines (**Figure 2.10J, K**). It is also noteworthy that highly dosage-sensitive phenotypes, like the A5 phenotype of *Abd-B* heterozygotes will reduce the potential to observe a strong effect in this assay. Indeed, the sensitivity of A5 segment pigmentation to *Abd-B* dosage in *D. yakuba* (**Figure 3.1)** highlights how the changes I have detected at *Abd-B* can significantly impact its pigmentation phenotype. Collectively, my results indicate that the strength of mutant phenotypes and the sensitivity of the tested genetic background should figure into interpretations derived from reciprocal hemizygosity testing. Concomitantly, my findings from reciprocal hemizygosity assays further suggested additional epistatic interactions among loci in the *D. yakuba* and *D. santomea* system.

3.4 A LINK BETWEEN MACRO- AND MICROEVOLUTION OF ANIMAL BODY PLANS

A recurring debate in evolutionary biology centers on whether macroevolutionary patterns of divergence represent the gradual accumulation of multiple microevolutionary changes, or are caused by the alteration of single loci with large effects (Coyne, 2006; Davidson and Erwin, 2006; Dobzhansky, 1937; Goldschmidt, 1940; Stern, 2000). This debate persists even today, due to several technical barriers (**Table 3.1**). Because macroevolutionary divergence exists between distantly related taxa, we are generally unable to use unbiased approaches such as QTL mapping to detect the causative changes that underlie macroevolutionary phenotypes. Thus, for macroevolutionary phenotypes, one is left with the candidate gene approach as the main method to identify causative mechanisms (**Table 3.1**). While genomic approaches such as RNA-seq offer an unbiased way to identify expression differences genomewide (**Table 3.1**), they provide only correlational evidence. Reporter assays, which represent a key test that can implicate *cis* regulatory changes underlying expression divergence face technical challenges at macroevolutionary depths as well. The high level of sequence divergence in non-coding regions can make it difficult to find orthologous regulatory regions among distantly related taxa (Rebeiz et

al., 2015). Furthermore, over long evolutionary distances, the landscape of transcription factors can diverge such that regulatory sequences become incompatible, and fail to activate their expression when transplanted from one species to another (Arnosti, 2003). This hampers our ability to use this approach to identify causal changes in regulation between distant taxa (**Table 3.1**).

Table 3.1 Comparisons of the feasibilities of the approaches applied between micro- and macroevolutionary studies

Approaches	Microevolution	Macroevolution
Analysis of Gene Expression	✔	✔
Reporter and Transgenic Assay	✔	✔ (*difficulties to arise due to *cis* / *trans* compatibility)
Genetic Mapping	✔	✔ (*only some crossable species)
Genomic Approaches (RNA seq…)	✔	✔

For body plan level traits, Hox genes represent excellent candidates, as their expression changes have long been found to correlate with obvious body plan shifts among vertebrates (Burke et al., 1995; Cohn and Tickle, 1999; Di-Poi et al., 2010), insects (Khila et al., 2009; Mahfooz et al., 2004; Rogers et al., 1997; Warren et al., 1994) and other arthropods (Averof and Patel, 1997; Martin et al., 2016). Given the important roles they play, and the dramatic homeotic transformations they produce in laboratory mutants, they have been posited to cause saltational transitions with huge effects in nature (Carroll, 1995; Gellon and McGinnis, 1998). However, only

a small number of studies have identified changes to Hox genes themselves, including coding changes (Ronshaugen et al., 2002; Galant and Carroll, 2002) and implications of non-coding *cis*-changes (Belting et al., 1998; Stern, 1998).

The comparison of *Ubx* between insects, crustaceans and onychophorans, which are separated by hundreds of million years revealed evidence for protein-coding changes in this Hox gene (Ronshaugen et al., 2002; Galant and Carroll, 2002). In both crustaceans and onychophorans, thoracic segments with normal appendages nevertheless express Ubx (Ronshaugen et al., 2002; Galant and Carroll, 2002; Averof and Patel, 1997) while in insects, such as *Drosophila*, *Ubx* greatly represses appendage development in the posterior thoracic segment and abdominal segments (Lewis, 1978; Casanova et al., 1985). A comparative analysis of the Ubx amino acid coding sequence function revealed a causative change in a C-terminal sequence outside of the homeodomain, a sequence that is responsible for decreasing limb repression function in crustaceans and onychophorans. Chimeric proteins which exchanged this C-terminal region provided species-specific suppression of limb formation in *Drosophila.* This demonstrated how in insects, the Ubx C-termini have evolved to form a poly-alanine motif which became unable to inhibit the limb repression function of Ubx (Ronshaugen et al., 2002; Galant and Carroll, 2002).

The above example suggests that evolution of the protein-coding regions of Hox genes could drive body plan evolution, especially among distantly related species. Also, a vast majority of cases show correlations of gene expression differences with massive shifts in morphology, which were implicated to be caused by *cis*-changes in Hox genes. Importantly, these may arise without phenotype-causing variation in the differentially expressed Hox gene. In the above case of *Ubx*, its role in suppressing the *Drosophila* appendage (haltere) formation has been reported to involve more than five downstream genes across different levels of the wing patterning network and each of them are independently regulated by Ubx (Weatherbee et al., 1998). On one hand, mutations in individual genes are not sufficient to transform the appendage morphology

completely, providing a way to protect body segmental identities from deleterious alterations. On the other hand, due to epistatic interactions among different levels of the network, it is difficult to predict the contribution of a Hox gene to the body plan phenotype. By focusing on one gene at a time, the majority of Hox-based studies of macroevolutionary divergence likely underestimate the true genetic complexity of evolutionary change.

The participation of Hox genes in body plan evolution also raises the question of how such highly pleiotropic genes can be modified during evolution. Do their changes represent saltational steps that are highly pleiotropic? Or are they gradually modified during evolution? In this work, I localized causative *cis*-changes to the Hox gene *Abd-B*, which provided insights into how this might occur more broadly. In this case, although the initiator element IAB5 is required for pigmentation, it is not sufficient to drive posterior body segment activity during pupal stages in which pigments form. The entire 15 kilobase *iab-5* region was required to recapitulate the full extent of Abd-B expression through pupal development, suggesting the existence of one or more maintenance elements which act in *cis* to IAB5. Through chimeric reporter gene constructs, I inferred the causative changes to reside in a maintenance element that altered how long Abd-B expression is maintained in the abdomen. Hence, the mutations I identified affect a temporally narrow window of development, causing a highly pleiotropic gene to have a limited effect on its phenotype.

Abd-B is required for multiple aspects of pupal abdomen morphology. In *D. yakuba* individuals homozygous for the IAB5 deletion, changes in pigmentation are accompanied by differences in the distribution of small hairs known as trichomes. Homozygote pupae exhibit loss of Abd-B expression in some patches of the A6 tergite (**Figure 3.3D**), perfectly matching the location of ectopic trichomes (**Figure 3.3E**). This phenotype contrasts with the lack of differences in trichome distribution between *D. yakuba* and *D. santomea* (Gompel and Carroll, 2003). Therefore, the temporally restricted changes I have found at *Abd-B* are able to circumvent this pleiotropic effect on trichomes. It is thus quite possible that changes similar to the ones I have

identified at Abd-B represent the initial stages of body plan divergence in which subtle changes to Hox genes are first made in temporally narrow windows that perhaps widen over time.

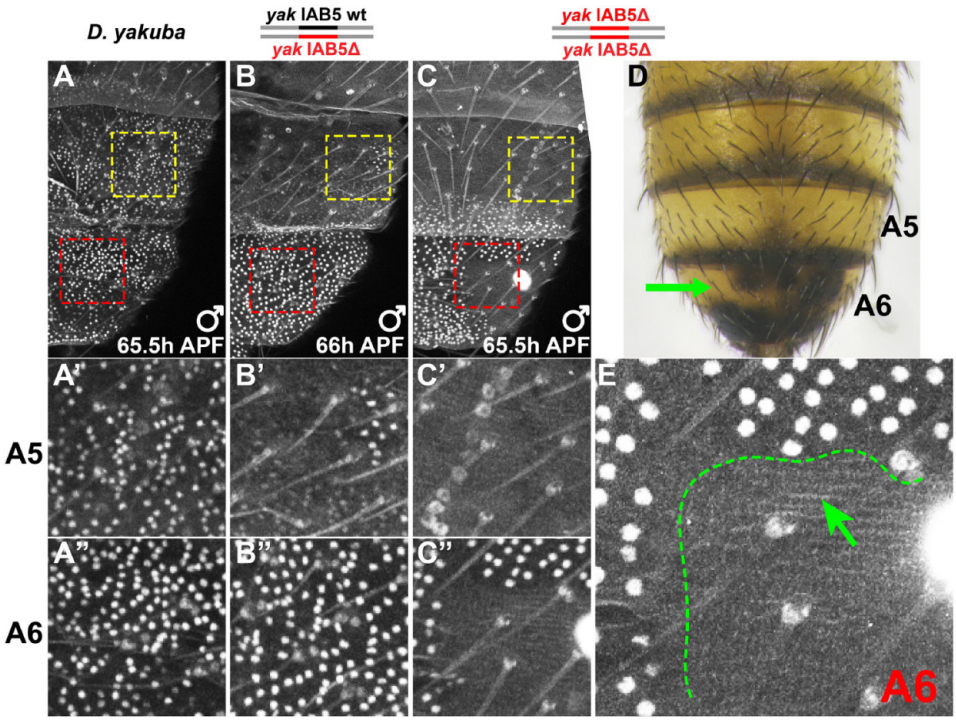

Figure 3.3 Pleiotropic effects of the initiator element IAB5. In addition to affecting abdominal pigmentation formation, the initiator element IAB5 is required to repress trichome formation by promoting Abd-B expression. (A-C) Abd-B expression in *D. yakuba* (A), a *D. yakuba* IAB5 deletion heterozygote (B), and a *D. yakuba* IAB5 deletion homozygous animal (C). The heterozygote IAB5 deletion animal shows partial decreases of Abd-B expression just in A5 (B), while animals homozygous for the deletion lacks Abd-B expression in the entire A5 and parts of the A6 segment (C). (A'-C', A"-C") Closeup images of the areas outlined in (A-C), showing epithelial cell expression. (D) Adult phenotype of a *D. yakuba* abdomen homozygous for the IAB5 deletion. Green arrow points to an area that lacks pigmentation in A6. (E) Higher magnification image of the area outlined by red in (C), showing that trichome pattern (green arrow, small gray lines) oppositely correlates with the Abd-B expression (outlined by green dashed line).

3.5 CONCLUDING REMARKS AND FUTURE DIRECTIONS

Using *Drosophila* abdominal pigmentation as a model, my thesis work uncovered how a body plan shift was built during evolution through the gradual accumulation of evolutionarily relevant *cis* regulatory changes throughout a whole Hox regulated gene network, including the top-tier transcription factors and terminal structural genes. Specifically, I have highlighted the CRE changes in five genes (**Figure 3.4**). The Hox transcription factor *Abd-B* has evolved *cis* change(s) in its maintenance element, leading to tissue specific loss of expression in late pupal developmental stages in *D. santomea*. Another transcription factor gene *pdm3* exhibited a gain of expression that enhanced its pigment-suppressing role. Three network terminal structural genes, *yellow*, *tan* and *ebony* have also evolved changes in their CREs that contribute to the *D. santomea* phenotype. The evolutionary mechanisms of *tan* and *ebony* demonstrate a parallel pattern of inactivation—*tan*, a dark pigment-promoting gene, has lost an activating CRE required for abdominal expression, while a transposon insertion inactivated a silencer element in the the light color-promoting gene *ebony*, leading to the expansion of expression in *D. santomea*. The combined effects of *tan* and *ebony* changes are to decrease dark pigmentation formation in *D. santomea*. Though the differences in *yellow* expression is mostly due to upstream changes at *Abd-B* and *pdm3*, *yellow* itself has also evolved small changes in its CRE to generate a subtle effect on the phenotype.

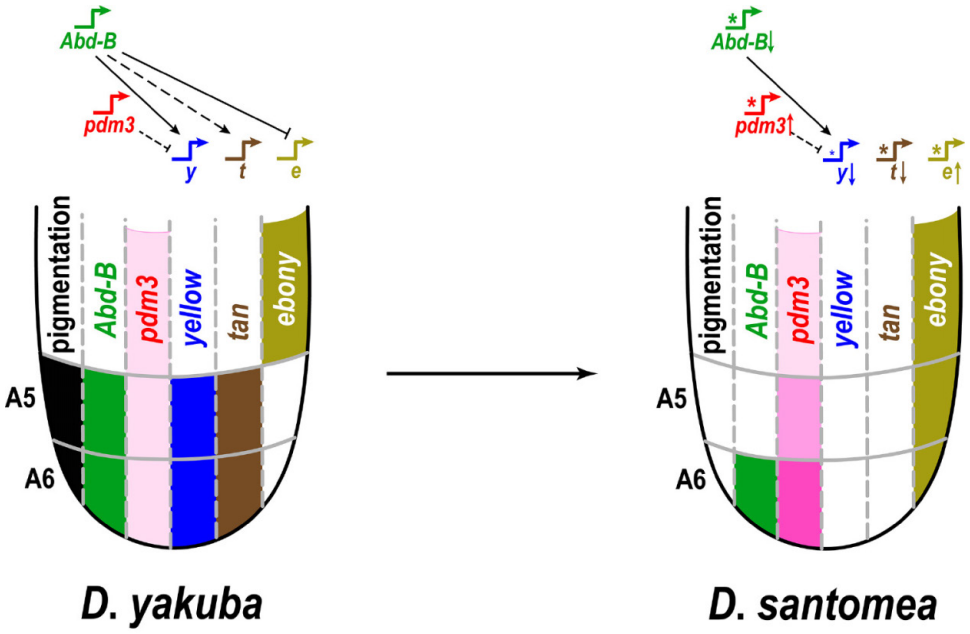

Figure 3.4 A microevolutionary portrait of body plan evolution. In the model, black color indicates pigmentation pattern, while other colors show gene expression in the abdominal segments. The networks above the abdomens show the hierarchical relationships among genes studied here, and indicate how they were are altered in *D. santomea*. The stars represent *cis*-regulatory changes, and arrows after gene name indicate up or down regulation. The long arrows or bars show activating and repressing interactions, respectively. Solid and dashed lines indicate direct and indirect (or unknown) interactions, respectively.

Now that we have a better understanding of the genes underlying the body plan shift between *D. yakuba* and *D. santomea*, a promising future direction will focus on how these contributing effects accumulated in the *D. santomea* population. To achieve that, we must first locate the exact changes in the regulatory elements of these altered genes, such as the causative change(s) in the maintenance element of *iab-5*, or the regulatory region of *pdm3*. Such experiments will allow us to survey *D. yakuba* and *D. santomea* populations to see whether the causative changes arose as new mutations or were likely standing genetic variants in the *D.*

yakuba/santomea ancestral population. Identification of the precise causative substitutions will also allow us to use CRISPR/Cas9 to transplant the evolutionary relevant changes from one species to another, and measure the degree of phenotypic change. If replacement of all the regions altered in these five loci can cause the phenotypic conversion of *D. yakuba* to the *D. santomea* pigmentation pattern, this example will fully confirm my complete dissection of a polygenic trait. Finally, it is worth considering the network depicted in **Figure 3.4**. Currently, this network is drawn with many fewer connections than actually exist in reality. As future work identifies causative substitutions in this system, my hope is that these mutations will be connected to their biological function in terms of which transcription factors differentially bind the sequences altered by phenotype-causing changes. This will inform not only the evolution of this trait, but also the endogenous function of these CREs, which include understudied classes of elements, such as silencers and maintenance elements.

Moving beyond the *D. yakuba/D. santomea* phenotypic difference, abdominal pigmentation is a promising system to look at macroevolutionary events further in the past. My observation that differences in Abd-B expression exist in *D. santomea* raises the exciting possibility that changes exist among more distantly related taxa in this clade. In a survey of the *melanogaster* subgroup species, I observed that the A5 expression of Abd-B is quite variable, suggesting repeated changes to Abd-B in this lineage (**Figure 3.5**). Because the gene regulatory network of the pigmentation pathway is so well characterized, this provides a near-comprehensive list of candidate genes which are required to study differences that exist beyond crossable species. Thus, many opportunities exist to explore the origins and diversification of this network with candidate gene approaches. As technologies for gene editing continue to improve, *Drosophila* pigmentation will endure as a premiere system in which to connect genetic variation to its phenotypic consequences, hopefully generating similarly complete views of phenotypic divergence at further taxonomic depths by transplanting large numbers of genes between taxa to

recapitulate complex macroevolutionary traits. Hence, my hope is that this work is the beginning of a widening bridge between microevolution and macroevolution.

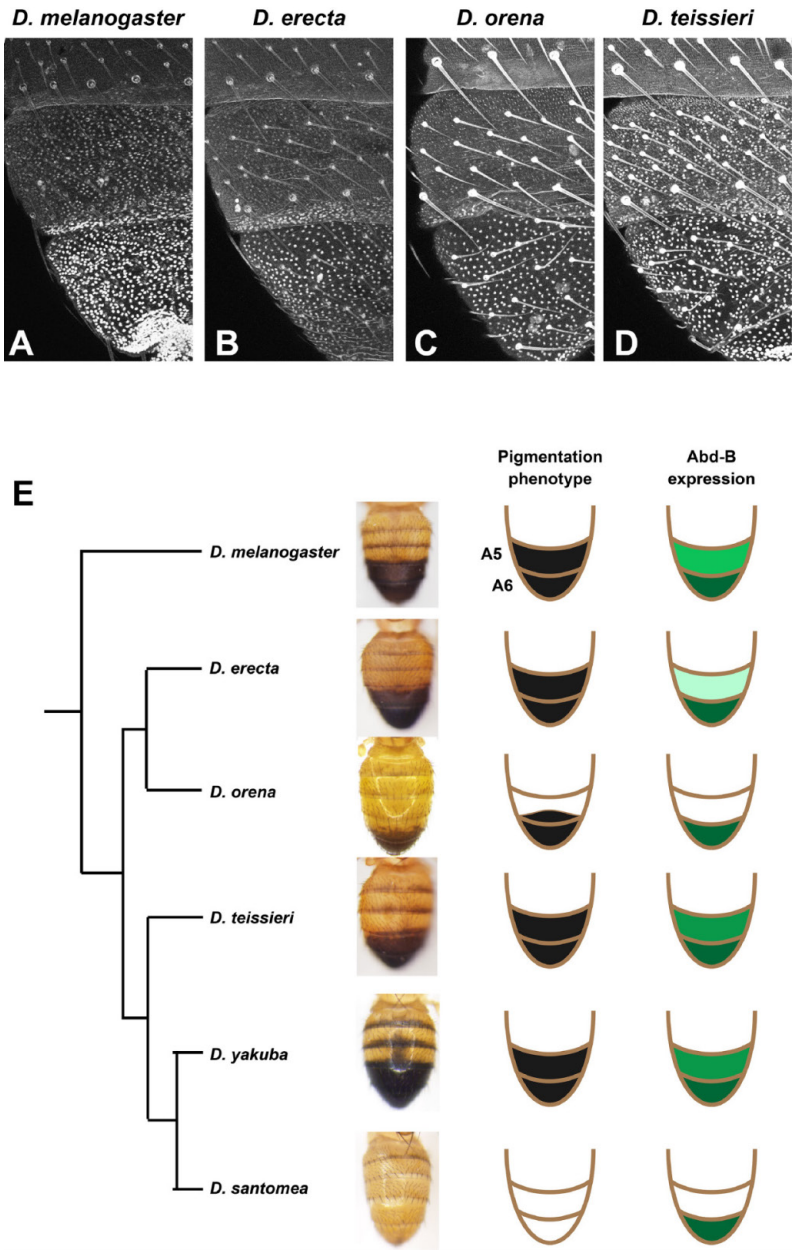

Figure 3.5 Variable expression of Abd-B the *melanogaster* subgroup. (A-D) Abd-B expression patterns in late male pupal abdomens of (A) *D. melanogaster*, (B) *D. erecta*, (C) *D. orena* and (D) *D. teissieri*. Especially in their A5 tergites, each species demonstrates different levels of expression. (E) Phylogeny of the *melanogaster* subgroup, showing adult phenotypes and a diagrammatic summary of pigmentation and Abd-B expression phenotypes. The darkness of green shading reflects relative expression levels.

APPENDIX: KEY RESOURCES, EXPERIMENTAL MODELS AND METHOD DETAILS

A.1 KEY RESOURCES AND PRIMERS TABLES

Table A1 Key resources.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-ABD-B homeobox protein	DSHB	Cat# 1A2E9
Mouse anti-Ubx 5' exon, stock concentration 53μg/mL	DSHB	FP3.38
Guinea pigs anti-Pdm3 protein	Gift from Cheng-Ting Chien Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan	reference DOI 10.1002/dneu.22003
Anti-ABD-A	Gift from Brian Gebelein	(Li-Kroeger et al., 2008)
Anti-Digoxigenin-AP Fab fragments	Roche Diagnostics GmbH Mannheim, Germany	Cat# LOT 14608124 REF 11093274910
Alexa Fluor 488 donkey anti-mouse IgG (H+L)	Life technologies	REF A21202 LOT 1423052
Alexa Fluor 594 goat anti-guinea pig IgG (H+L)	Invitrogen Thermo Fisher Scientific Life Technologies	REF A11076 LOT 1848493
Bacterial and Virus Strains		
DH5 alpha		
Stella cells from Infusion cloning kit	Clontech (EV) TAKARA	
Chemicals, Peptides, and Recombinant Proteins		
paraformaldehyde 16% aqueous solution (Methanol free)	Electron Microscopy Sciences	Cat# 15710
Formamide, 500mL	Fisher Scientific	BP227500
Glycerol, 4L	EMD Millipore Corporation	MGX01855
Heptane, 500mL	Fisher Scientific	BP1115-500
NBT/BCIP color development substrate	Promega	REF# S3771

Table A1 (continued).

Phusion Flash High-Fidelity PCR Master Mix	Fisher Thermo Scientific	Cat# F-548L
Proteinase K	Fisher	BP1700-50
salmon sperm DNA sodium salt	Fisher	50-591-966
Dig RNA labeling mix 10X	Roche Diagnostics GmbH	Cat# 17109821
Halocarbon Oil 27	Sigma-Aldrich	H8773
Halocarbon Oil 700	Sigma-Aldrich	H8898
Heparin sodium salt from porcine intestinal mucosa	Sigma-Aldrich	H4784
Tween-20	Fisher Scientific	BP337-500
Triton-X-100 electrophoresis grade	Fisher Scientific	BP151-500
T7 RNA polymerase	New England Bio Labs	Cat# M0251S
RNase inhibitor, Murine	New England Bio Labs	Cat# M0314S
Asc I	New England Bio Labs	REF# R0558L
Sbf I	New England Bio Labs	REF# R3642L
Spe I	New England Bio Labs	REF# R0133L
T4 DNA ligase	New England Bio Labs	REF# M0202L
Critical Commercial Assays		
QIAprep Spin MiniPrep Kit	Qiagen GmbH	Cat# 27104
QIAquick Gel Extraction Kit	Qiagen GmbH	Cat# 28706
QIAGEN Plasmid *Plus*Midi Kit	Qiagen GmbH	Cat# 12945
In-Fusion HD Cloning Plus	Clontech (EV) TAKARA	Cat# 638909
Experimental Models: Organisms/Strains		
yellow white: *yw*;+;+		
Drosophila yakuba	*Drosophila* Species Stock Center	
Drosophila santomea	*Drosophila* Species Stock Center	
Drosophila melanogaster: M[vas-int.Dm]ZH-2A, M[3xP3-RFP.attP]ZH-51D	Rainbow Transgenic Flies	BLS# 24483
Drosophila melanogaster: M[vas-int.Dm]ZH-2A, P[CaryP]attP2	Rainbow Transgenic Flies	BLS# R8622
Drosophila melanogaster: P[nos-phiC31\int.NLS]X, P[CaryP]attP2	Rainbow Transgenic Flies	BLS# 25710
Oligonucleotides		
Primers for probe synthesis see **Table A2**	This thesis	
Primers for cloning transgenic GFP reporter constructs see **Table A3**	This thesis	
Primers for infusion cloning or overlapping PCR cloning of GFP reporter constructs see **Table A4**	This thesis	
Primers for sequencing or testing see **Table A5**	This thesis	
Guiding RNAs sequences for generating CRISPR flies see **Table A6**	This thesis	

Table A1 (continued).

Software and Algorithms		
InsituPro VIS liquid handling system (*insitu* robot)	Intavis Bioanalytical Instruments	N/A
Genepalette		(Rebeiz and Posakony, 2004)

Table A2 Primers for *in situ* hybridization.

Tested Gene	Forward Primer	Reverse Primer	Notes
yak/san yellow	TTYGCCGTMTCCACGAGGAT	taatacgactcactataggAKGCCGTTGTGCTGGTTGAA	A T7 promoter was appended to the reverse primer.
yak/san ebony	AGCTATCGCCAGATGAACGAG	taatacgactcactataggGTCTTGAAAACGCTCACCGTCTC	A T7 promoter was appended to the reverse primer.
yak/san tan	GACGGAGACCCTGAATCACTAC	taatacgactcactataggGTTTTGCCGCTGCGCAAGAGCTC	A T7 promoter was appended to the reverse primer.
GFP	atttaggtgacactatagaCCACCATGGTGAGCAAGGGCGAGG	taatacgactcactataggTTAGCGTCTTCGTTCACTGCTGCG	A T7 promoter was appended to the reverse primer, while an SP6 promoter was appended to the forward primer.

Table A3 Primers for cloning transgenic GFP reporter constructs

Construct	Forward Primer	Reverse Primer	Restriction Sites
yak/*san* IAB5 core element	ATAGATCTGGTCTAGAGCCCGG GCGAATTCGCCggcgcgccCAATT GCCCAGGTATCTCCA	GGATCCGCTAGCTTCCGCGGT TGCGATCGCTTcctgcaggTTCCA CTTCCGAACTTGGTC	*Asc I*/*Sbf I*
upstream region of *yak*/*san ebony* upstream + intron	TTCCGggcgcgccGAGCAACCCTT TTTATAAGCGATG	TTGCCcctgcaggCCTGCTCTTAM AGCCSCTGCAATTAC	*Asc I*/*Sbf I*
intron region of *yak*/*san ebony* upstream + intron	CATCAATGTATCTTAactagtCTGC GAGCGCCGTTTACAAGTACA	CACACTTATTACGTGactagtAGC TGCTGCTCCTCGAAGATGCGG	*Spe I*
yak/*san yellow* wing+body element	TTCCGggcgcgccCTCCTCCATGG TGGTGGAACTA	TTGCCcctgcaggACGACTGGTG GCCATAATAAGTC	*Asc I*/*Sbf I*
yak/*san yellow* body element	TTCCGggcgcgccGCTTTCCGCCC AAGTTGAAGTG	TTGCCcctgcaggCGGGTAATCA GGTGGCTTATGC	*Asc I*/*Sbf I*

Table A4 Primers for infusion cloning or overlap extension PCR cloning of GFP reporter constructs

Name	Primer	Notes
iab-5-F-*Asc I*	GGGCGAATTCGCCggcgcgccCRTTTTCCGTTTTATTGCGA	For amplifying *yak/san iab-5* fragment 3
iab-5-R-*Sbf I*	TTGCGATCGCTTcctgcaggCGGCCGATGAAAGCAGTCCGCCAG	For amplifying *yak/san iab-5* fragment 1
yak-iab-5-int-F3	CATGCTGCCATATTGCCAGAAC	For amplifying *yak iab-5* fragment 1
yak-iab-5-int-R3	GTTCTGGCAATATGGCAGCATG	For amplifying *yak iab-5* fragment 2
iab-5-whole MidF	GATGAGATTCAAGTGGCTGCTTTC	For amplifying *yak/san iab-5* fragment 2
iab-5-whole MidR	GAAAGCAGCCACTTGAATCTCATC	For amplifying *yak/san iab-5* fragment 3
san-iab-5-intF6	GTCAATTAGCTGGTGCCAGTGTG	For amplifying *san iab-5* fragment 1
san-iab-5-intR6	CACACTGGCACCAGCTAATTGAC	For amplifying *san iab-5* fragment 2
*san*EndSeqR1	CTGACGAAATTCCGACGGGAG	For amplifying *san iab-5* pre-fragment 1
yaksan-iab-5-int-F5	GTCTTCCATGTCTACGCCTGTTTG	For amplifying *san iab-5* pre-fragment 1
YS*eb*US-chim-F	CTAGAGCCCGGGCGAATTCGCCggcgccGATAAGGATTAGTWATATATGRRC	External forward primer for overlap extension PCR to clone the *yak ebony* Full + TE construct upstream region (Restriction site *Asc I*)
YS*eb*US-chim-R	TCCGCGGTTGCGATCGCTTcctgcaggCCTGCTCTTACAGCCGCTGCAATTAC	External reverse primer for overlap extension PCR to clone the *yak ebony* Full + TE construct upstream region (Restriction site *Sbf I*)
yaksan-chim-hltr-F1	GCGCTATTAAAGGTGTACTTGCTCG	Internal primer for overlap extension PCR to clone the *yak ebony* Full + TE construct upstream region
yaksan-chim-hltr-R1	CGAGCAAGTACACCTTTAATAGCGC	Internal primer for overlap extension PCR to clone the *yak ebony* Full + TE construct upstream region
yaksan-chim-hltr-F2-3	CAATTGAAATGATAAATCCGCTCATTATTCTTGAACTCAC	Internal primer for overlap extension PCR to clone the *yak ebony* Full + TE construct upstream region
yaksan-chim-hltr-R2-3	GTGAGTTCAAGAATAATGAGCGGATTTATCATTTCAATTG	Internal primer for overlap extension PCR to clone the *yak ebony* Full + TE construct upstream region

Table A5 Primers for sequencing or testing.

Name	Primer	Notes
yaksan-iab-5-int-R4	CAGCAGGTGAGTCCATTAAGTG	Sequencing through the junction between fragment 1 and 2 of *yak*/*san iab-5* constructs
*iab5*_seq_junc2F	GTGCAGCGTGATGTTCATTACAC	Sequencing through the junction between fragment 2 and 3 of *yak*/*san iab-5* constructs
S3aF2	CACATGTGCAAGAGAACCCAGTG	Sequencing the upstream region of the regulatory sequence of GFP reporter construct in the S3aG vector
S3aR2	CTGCGCTTGTTTATTTGCTTAGC	Sequencing the downstream region of the regulatory sequence of GFP reporter construct in the S3aG vector
san-eb-US-tps-F2	GAGAACATTGTTGCCGACAAGC	Forward primer to amplify the region containing the *santomea* specific transposable element. The primer works in *yakuba* and *santomea*.
*san_eb*_US_tps_R2	TGCCAGCCGTCATGTTGTGCTTC	Reverse primer to amplify the region containing the *santomea* specific transposable element. The primer works in *yakuba* and *santomea*.
yaksan-chim-hltr-R2-2	CTCAATGTGGTCCCATTTGCATTCG	Primer to sequence through the transposable element.
S3aG-*Spe*-F2	CGCCGTCAACGGAGCCGAAGT	Primer to sequence *ebony* intron region placed downstream of GFP in the S3AG vector

Table A6 Guide RNAs for generating CRISPR flies.

Gene or Regulatory Region to Mutate	gRNA Sequence	Notes
yellow-gRNA9	GGCCAACGGACUGAAGUACAGGG	
ebony-gRNA5	GGUUCGCCCGCUCGUUCAUCUGG	co-injected
ebony-gRNA6	GGUCUCGGCCACCAGGAGACGGG	
tan-gRNA2	GGAGUACAGGGAGAUGGUCCUGG	cut left site (Double check the sequence)
tan-gRNA1	AAUCACUCACGCAAUUCUUCUGG	cut right site (Double check the sequence)
IAB5-gRNA1	GGUGCGUUUCCAUUUUCCCUWGG	cut left site
IAB5-gRNA2	GGWUCAAUCGGYUUAUUGAUGGG	cut right site
pdm3-gRNA1-*yak*	GAUGAAAUGGAGAUCACAGACGG	cut left site
pdm3-gRNA4-*yak*	ACGUGAUCCUUGUUGUCGAACGG	cut right site
pdm3-gRNA1-*san*	GGAGUUCUACAAGAACCUGGCGG	cut left site (Note that *san* CRISPR region is 9bp downstream of *yak* CRISPR region in the alignment.)
pdm3-gRNA2-*san*	GCGUUUGCCGCCCAUCUGAACGG	cut right site

A.2 EXPERIMENTAL MODEL AND SUBJECT DETAILS — *DROSOPHILA* STRAINS AND CULTURE

Stocks were maintained at room temperature on standard cornmeal agar media. *D. yakuba Ivory Coast*, *D. yakuba Jess* (14021-0261-01), *D. santomea* STO.4 (14021-0271.00) and *D. santomea* STO CAGO1482 were obtained from the Drosophila Species Stock Center (http://blogs.cornell.edu/drosophila/). A *melanogaster yellow white* (*yw*) strain that was isogenized for eight generations was used for crosses to normalize the backgrounds of GFP reporter transgenes. A total of 78 *D. yakuba* and 128 *D. santomea* iso-female lines were used to test the ~500 bp transposable element insertion in the *D. santomea ebony* male repression element by PCR screening (work from Clair Han).

A.3 EXPERIMENTAL METHODS

A.3.1 *in situ* hybridization

in situ hybridization was performed as described in (Jeong et al., 2008) with small modifications. In brief, pupal samples were aged to differing extents for each probe, dissected in cold PBS, and fixed in PBS containing 4% paraformaldehyde (E.M.S. Scientific), .1% Triton X-100. PCR was performed to generate RNA probe templates that had a T7 promoter appended through primer design. Templates were sequenced verified using a T7 sequencing primer. See **Table A2** for the list of probe primers. Digoxigenin-labeled probes were generated using a 10X Dig labeling mix (Roche Diagnostics) and T7 RNA polymerase (Promega). Dissected samples were probed using an *in situ* hybridization robot (Intavis). Because *D. yakuba* and *D. santomea* are very closely related, spatial differences in pigmentation gene expression could reliably be detected with the same probe made from one species. To ensure the maximal ability to detect expression, I used probes from the species predicted to have the lowest expression level. Therefore, the *yellow*

probe was made from *D. santomea*, while the *ebony* probe was generated from *D. yakuba*. To compare the expression patterns and levels among *D. yakuba*, *D. santomea* and the introgression lines, samples were treated with the same concentration of probe under identical experimental conditions, including staining time. For GFP *in situ* hybridization, the probe template was amplified from the transgenic S3aG vector used in reporter assays.

A.3.2 GFP transgenic reporter assays

D. yakuba and *D. santomea ebony* and the 1 kb initiator element IAB5 reporter constructs were PCR amplified, and cloned by enzyme digestion into the S3AG vector (Williams et al., 2008), which contains a basal promoter driving GFP, flanked by SF1 and gypsy insulators. Insertion of the *ebony* intron downstream of GFP was performed using the Infusion cloning kit (Clontech), inserting intron sequences into a *Spe I* site downstream of GFP. The full *iab-5* reporter constructs were generated by the assembly of multiple fragments by the Infusion method (Clontech), inserting between *Asc I* and *Sbf I* sites upstream of GFP (**Figure 4.1**). See **Table A4** for primers used to clone regulatory constructs. Primers were designed using the GenePalette Software tool (Rebeiz and Posakony, 2004; Smith et al., 2017). Restriction sites (*Asc I* and *Sbf I*) were appended to primers (Integrated DNA Technologies) for insertion into the S3AG multicloning site. Transformant lines were generated by phiC31 mediated site specific recombination into the 51D insertion site on the second chromosome (Bischof et al., 2007). In the case of *iab-5* reporter constructs, transgenes were also inserted into the attP2 site on the third chromosome (Groth et al., 2004). Transgenic animals were mounted on slides in halocarbon oil and imaged on an Olympus Fluoview 1000 confocal microscope. For *iab-5* reporter constructs, flies were mounted at 94-95h after pupal formation (APF) in halocarbon oil 27 and imaged, or dissected for *in situ* hybridization to detect GFP mRNA. For *yellow* reporter constructs, flies were mounted at 91h after pupal formation in halocarbon oil 27 and imaged. For *ebony* reporters, samples were aged for 6

hours after eclosion before mounting in halocarbon oil 700 and imaging. Samples were imaged with standard settings in which the brightest samples were not saturated.

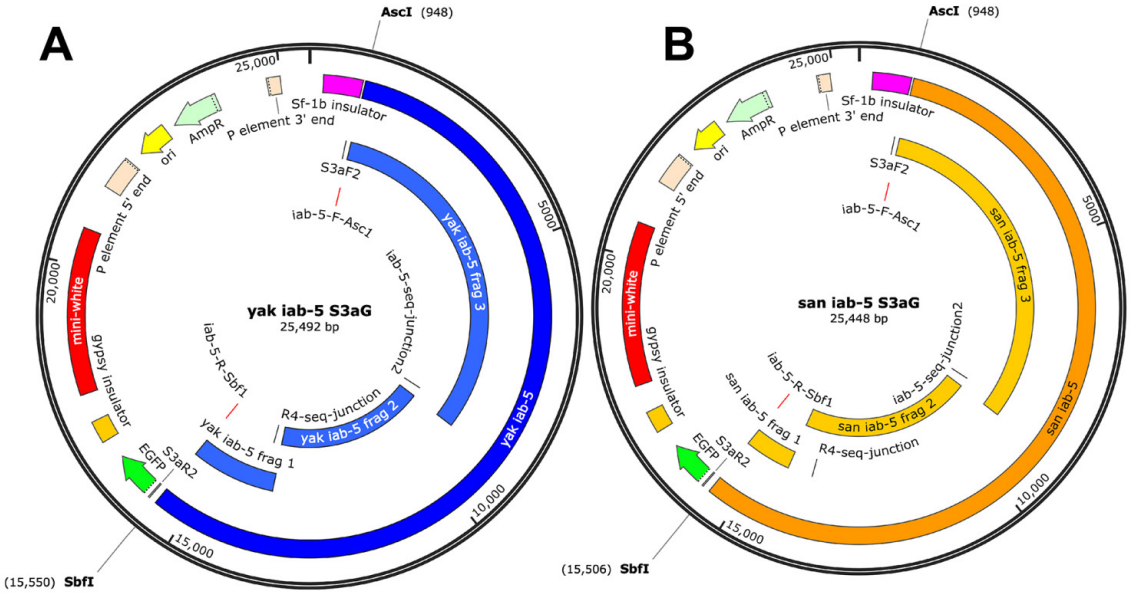

Figure A1 Schematics detailing the assembly of *D. yakuba* and *D. santomea* full *iab-5* GFP reporter constructs by infusion cloning. Primers sequences are presented in Tables A4 and A5.

A.3.3 Imaging of GFP reporters in embryos

The *D. yakuba* and *D. santomea* IAB5 (1kb) and *iab-5* (~15kb) transgenic GFP reporter embryos were collected at Stage 11 from Grape Agar plates, treated with fresh 50% bleach for 3 minutes, followed by rinsing with deionized H_2O containing 0.1% Triton. Embryos were then transferred into a mixture of 3mL heptane and 3mL PBS fix (PBS fix: PBS containing 4% paraformaldehyde (E.M.S. Scientific)) and were rotated on an orbital shaker set to 200-250 rpm at room temperature for 20 minutes. The bottom layer (PBS fix) was removed, and 3mL Methanol was added, followed by vigorous shaking by hand for 1 minute. The devitallenized embryos were transferred into a 1.5 mL Eppendorf tube and the supernatant was removed and replaced with PBT (PBS containing 0.1% Triton). Methanol exposure was minimized during the process to reduce the degradation of GFP. The embryos were suspended in glycerol mounting solution (80% glycerol, 0.1M Tris, pH 8.0) and mounted on slides. Imaging was performed with on an Olympus Fluoview 1000 confocal microscope.

A.3.4 CRISPR/Cas9 induced transgenic mutant flies

The null mutations of *ebony* and *yellow* were generated by CRISPR/Cas9 mediated mutagenesis by non-homologous end joining. Mutant alleles were identified by screening for mutant phenotypes. The IAB5, *pdm3*, and *tan* mutations were induced by CRISPR/Cas9 mediated homologous recombination (Yu et al., 2013). In each case, native genomic DNA was replaced with *3xP3*::*DsRed* to provide a dominant marker for identification of mutant alleles and stock maintenance. The guide RNA sequences are listed in **Table A6**.

A.3.5 Reciprocal hemizygosity tests

Following the scheme that is illustrated in **Figure 2.4**, the reciprocal hemizygosity test (RHT) compares the progeny of two separate crosses. The first group is composed of offspring from *D. yakuba* wildtype to *D. santomea* CRISPR mutants (**Figure 2.4E**). The second group has offspring from *D. santomea* wildtype crossed to *D. yakuba* CRISPR mutants (**Figure 2.4F**). For *tan*, IAB5 and *yellow*, I performed *D. yakuba*-mothered crosses, which can generate offspring fairly reliably. For *pdm3* and *ebony*, to decrease the strong influence of the *D. yakuba* X chromosome (containing *yellow* and *tan*), I performed *D. santomea*-mothered crosses, which are generally far less successful. For *ebony*, I also performed an RHT test using mutant introgression lines, which provides a more sensitive background to detect phenotypic differences among progeny. Instead of using *D. yakuba* as a parent, I used the Introgression IIIB line that contains a *D. yakuba* version of the *ebony* gene.

A.3.6 Imaging of adult *Drosophila* abdomens

The eclosed progeny from the reciprocal hemizygosity test crosses were directly transferred into new vials every one to two days to avoid the effects of CO_2 on the tanning process. After culturing at room temperature for 6 to 9 days, flies were selected on a CO_2 pad for imaging. Note that *ebony* and *yellow* mutant lines were homozygous and all progeny could be phenotyped. In IAB5, *pdm3* and *tan* CRISPR lines, mutant alleles were marked by *3xP3*::*DsRed* (RFP), which labels the eyes. However, only *tan* CRISPR lines were homozygous and the *D. santomea* IAB5 CRISPR line was completely balanced by an induced inversion in the bithorax complex that increased the haltere's size. Thus, the offspring from these RHT crosses needed to be selected for 3XP3-RFP

fluorescence. Abdomens were mounted on slides covered with double-sticky tape, and imaged using Leica M205C Stereo Microscope with a DFC425C camera. The Lecia Montage package was used to generate an extended focus brightfield image.

A.3.7 Immunohistochemistry

Immunohistochemistry of pupal abdominal epidermis was performed as previously described (Gompel and Carroll, 2003). For Abd-B, flies were dissected in cold PBS and fixed in PBS containing 4% paraformaldehyde (E.M.S. Scientific) and 0.1% Triton-X-100 (PBT-fix) at 63h (for *D. yakuba*) and 66-71h (*D. santomea* and introgression lines) APF respectively since *D. yakuba* and *D. santomea* develop at different rates. Abdominal, larval or embryonic samples were incubated at 4°C in 1:100 diluted primary antibody (Developmental Studies Hybridoma Bank, stock concentration 35μg/mL) in PBS containing 0.1% Triton-X-100 (PBT) overnight, followed by three 15 minute PBT washes at room temperature. The samples then were transferred into 1:500 diluted secondary antibody (Alexa Fluor 488 donkey anti-mouse IgG, Invitrogen Thermo Fisher Scientific Life Technologies, stock concentration 2mg/mL) in PBT and incubated at room temperature for 2.5h, followed by three washes in PBT, each for 15 minutes. Finally, the samples were mounted in glycerol mounting solution (80% glycerol, 0.1M Tris, pH 8.0) and imaged.

For the Abd-B immunostainings of the IAB5 reciprocal hemizygosity (RH) cross offspring, since the parental homozygous viable CRISPR lines are difficult to obtain, and the fluorescence during pupal stages is difficult to assess, I screened for animals carrying the IAB5 deletion based on the fluorescence of the 3xP3 cassette in the larval nervous system. Fluorescent larvae were allowed to grow on grape agar plates and were aged to 65 hAPF upon the formation of white pre-pupa to insure the proper staging of samples.

For stainings involving Pdm3, I incubated the samples in the primary antibody (Guinea pig anti-Pdm3, (Chen et al., 2012)) at a 1:100 dilution at 4°C for 36h (instead of overnight). The

secondary antibody (Alexa Fluor 594 goat anti-guinea pig IgG, Invitrogen Thermo Fisher Scientific Life Technologies, stock concentration 2mg/mL) was used at 1:500 dilution and incubated for 6h before washing.

For *abd-A*, we used a Guinea Pig anti-Abd-A (Li-Kroeger et al., 2008) at a 1:500 dilution, detected by Alexa Fluor 647 anti-guinea pig IgG (H+L) by Life Technologies (#A-21450) at 1:500 dilution. For stainings of Ubx, I incubated the samples in 1:100 diluted antibody (Mouse anti-Ubx 5' exon, Developmental Studies Hybridoma Bank, stock concentration 53μg/mL) overnight, detected by 1:500 diluted secondary antibody (Alexa Fluor 488 donkey anti-mouse IgG, Invitrogen Thermo Fisher Scientific Life Technologies, stock concentration 2mg/mL) incubation at room temperature for 2.5h.

A.3.8 Quantification and Statistical Analysis

A.3.8.1 Quantification of relative fluorescent intensity. Relative fluorescent intensity was measured based on the method previously described (Jeong et al., 2008; Rebeiz et al., 2009a) with minor modifications.

For *ebony* reporters, the light intensity (L) of A5 or A6 segment fluorescence was measured by the mean value of hemisegments outlined by the free hand selection tool in Adobe Photoshop. The background intensity (B) of A5 or A6 was measured as the average value of the intensities from three different small square patches from the midline tissue of the corresponding body segment that lacked epithelial cell or bristle fluorescence. The absolute light intensity (AL) of each segment was obtained from $L - B$. The relative percentage of fluorescent intensity of each segment was calculated as:

$$RF_i(\%) = \frac{L_i - B_i}{average\ value\ of\ AL\ (santomea\ A5)} \times 100(\%) \quad \textbf{Equation A1}$$

For quantifying the fluorescent intensity of *yellow* reporter fly abdomens, the absolute light intensity (AL) was calculated by measuring the light intensity (L) of A5 or A6 segments, subtracting the background intensity (B) from A4 intensity value from each sample. The relative percentage of fluorescent intensity of each segment was measured as:

$$RF_i(\%) = \frac{L_i - B_i}{average\ value\ of\ AL\ (yakuba\ A6)} \times 100(\%) \quad \textbf{Equation A2}$$

A.3.8.2Quantifications of *Drosophila* abdominal pigmentation. Images were converted into grayscale in Adobe Photoshop. The light value (L) was recorded as the mean value on a 0-255 scale, from free hand selection of the segment of interest. The segment intensity was obtained through:

$$segment\ intensity(\%) = \frac{255 - L}{255} \times 100(\%) \quad \textbf{Equation A3}$$

A two-tailed Student's t-test was used when comparing two sets of quantification data. The difference was defined to be significant when p value was less than 0.05. The number of stars in all plots indicate *p* values as follows: ****: *p*<0.0001, ***: *p*<0.001, **: *p*<0.01, *: *p*<0.05, ns: not significant. Box plots in all figures were drawn via the boxplot server (http://shiny.chemgrid.org/boxplotr/). Data points were plotted as jittered plots.

A.4 DATA AND SOFTWARE AVAILABILITY

Genepalette software for primer designing, visualizing and manipulating genomic sequences: http://www.genepalette.org/

BIBLIOGRAPHY

Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T.T., Mast, J., Sunayama-Morita, T., and Stern, D.L. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res *21*, 610-617.

Arakane, Y., Lomakin, J., Beeman, R.W., Muthukrishnan, S., Gehrke, S.H., Kanost, M.R., and Kramer, K.J. (2009). Molecular and functional analyses of amino acid decarboxylases involved in cuticle tanning in *Tribolium castaneum*. J Biol Chem *284*, 16584-16594.

Arnone, M.I., and Davidson, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development *124*, 1851-1864.

Arnone, M.I., Dmochowski, I.J., and Gache, C. (2004). Using reporter genes to study *cis*-regulatory elements. Methods Cell Biol *74*, 621-652.

Arnosti, D.N. (2003). Analysis and function of transcriptional regulatory elements: insights from *Drosophila*. Annu Rev Entomol *48*, 579-602.

Ashburner, M., Golic, K.G., and Hawley, R.S. (2005). *Drosophila*: a laboratory handbook, 2nd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

Averof, M., and Patel, N.H. (1997). Crustacean appendage evolution associated with changes in Hox gene expression. Nature *388*, 682-686.

Bachtrog, D., Thornton, K., Clark, A., and Andolfatto, P. (2006). Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. Evolution *60*, 292-302.

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729-740.

Bassett, A.R., Tibbit, C., Ponting, C.P., and Liu, J.L. (2013). Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep *4*, 220-228.

Belting, H.G., Shashikant, C.S., and Ruddle, F.H. (1998). Modification of expression and *cis*-regulation of *Hoxc8* in the evolution of diverged axial morphology. Proc Natl Acad Sci U S A *95*, 2355-2360.

Bischof, J., Maeda, R.K., Hediger, M., Karch, F., and Basler, K. (2007). An optimized transgenesis system for *Drosophila* using germ-line-specific phi C31 integrases. P Natl Acad Sci USA *104*, 3312-3317.

Blackwood, E.M., and Kadonaga, J.T. (1998). Going the distance: a current view of enhancer action. Science *281*, 60-63.

Bourne, G.H. (1969). The chimpanzee; a series of volumes on the chimpanzee (Basel ; New York: S. Karger).

Brakefield, P.M., Gates, J., Keys, D., Kesbeke, F., Wijngaarden, P.J., Montelro, A., French, V., and Carroll, S.B. (1996). Development, plasticity and evolution of butterfly eyespot patterns. Nature *384*, 236.

Bridges, C.B., and Morgan, T.H. (1923). The third-chromosome group of mutant characters of *Drosophila melanogaster* (Washington,: Carnegie Institution of Washington).

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

Budd Graham, E. (1999). Does evolution in body patterning genes drive morphological change—or vice versa? BioEssays *21*, 326-332.

Burke, A.C., Nelson, C.E., Morgan, B.A., and Tabin, C. (1995). Hox genes and the evolution of vertebrate axial morphology. Development *121*, 333-346.

Busturia, A., and Bienz, M. (1993). Silencers in *Abdominal-B*, a homeotic *Drosophila* gene. Embo J *12*, 1415-1425.

Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the *Drosophila Abd-B* gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. Development *128*, 2163-2173.

Camino, E.M., Butts, J.C., Ordway, A., Vellky, J.E., Rebeiz, M., and Williams, T.M. (2015). The evolutionary origination and diversification of a dimorphic gene regulatory network through parallel innovations in *cis* and *trans*. Plos Genet *11*.

Carbone, M.A., Llopart, A., DeAngelis, M., Coyne, J.A., and Mackay, T.F.C. (2005). Quantitative trait loci affecting the difference in pigmentation between *Drosophila yakuba* and *D. santomea*. Genetics *171*, 211-225.

Carroll, S.B. (1995). Homeotic genes and the evolution of arthropods and chordates. Nature *376*, 479-485.

Carroll, S.B. (2005). Evolution at two levels: on genes and form. Plos Biol *3*, e245.

Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: A genetic theory of morphological evolution. Cell *134*, 25-36.

Carroll, S.B., Gates, J., Keys, D.N., Paddock, S.W., Panganiban, G.E., Selegue, J.E., and Williams, J.A. (1994). Pattern formation and eyespot determination in butterfly wings. Science *265*, 109-114.

Carroll, S.B., Grenier, J.K., and Weatherbee, S.D. (2005). From DNA to diversity: molecular genetics and the evolution of animal design, 2nd edn (Malden, Mass. ; Oxford: Blackwell Pub.).

Casanova, J., Sánchez-Herrero, E., and Morata, G. (1985). Prothoracic transformation and functional structure of the Ultrabithorax gene of Drosophila, Cell *42*, 663-669.

Celniker, S.E., Sharma, S., Keelan, D.J., and Lewis, E.B. (1990). The molecular-genetics of the bithorax complex of *Drosophila*: *cis*-regulation in the *Abdominal-B* domain. Embo J *9*, 4277-4286.

Chan, S.K., and Mann, R.S. (1993). The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. Genes Dev *7*, 796-811.

Chan, Y.F., Marks, M.E., Jones, F.C., Villarreal, G., Jr., Shapiro, M.D., Brady, S.D., Southwick, A.M., Absher, D.M., Grimwood, J., Schmutz, J.*, et al.* (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. Science *327*, 302-305.

Chandler, C.H., Chari, S., and Dworkin, I. (2013). Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet *29*, 358-366.

Chauvet, S., Merabet, S., Bilder, D., Scott, M.P., Pradel, J., and Graba, Y. (2000). Distinct hox protein

sequences determine specificity in different tissues. Proc Natl Acad Sci U S A *97*, 4064-4069.

Chen, C.K., Chen, W.Y., and Chien, C.T. (2012). The POU-domain protein Pdm3 regulates axonal targeting of R neurons in the *Drosophila* ellipsoid body. Dev Neurobiol *72*, 1422-1432.

Cohn, M.J., and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. Nature *399*, 474-479.

Coyne, J.A. (2006). Comment on "Gene regulatory networks and the evolution of animal body plans". Science *313*, 761; author reply 761.

Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F.*, et al.* (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell *160*, 191-203.

D. Hoogland, R., Morris, D., and Tinbergen, N. (1956). The Spines of Sticklebacks (Gasterosteus and Pygosteus) as Means of Defence Against Predators (Perca and Esox), Vol 10.

Damen, W.G., Hausdorf, M., Seyfarth, E.A., and Tautz, D. (1998). A conserved mode of head segmentation in arthropods revealed by the expression pattern of Hox genes in a spider. Proc Natl Acad Sci U S A *95*, 10665-10670.

Davidson, E.H. (2006). The regulatory genome: gene regulatory networks in development and evolution (Burlington, MA ; San Diego: Academic).

Davidson, E.H., and Erwin, D.H. (2006). Gene regulatory networks and the evolution of animal body plans. Science *311*, 796-800.

Deutsch, J.S., and Mouchel-Vielh, E. (2003). Hox genes and the crustacean body plan. Bioessays *25*, 878-887.

Di-Poi, N., Montoya-Burgos, J.I., Miller, H., Pourquie, O., Milinkovitch, M.C., and Duboule, D. (2010). Changes in Hox genes' structure and function during the evolution of the squamate body plan. Nature *464*, 99-103.

Ding, Y., Berrocal, A., Morita, T., Longden, K.D., and Stern, D.L. (2016). Natural courtship song variation caused by an intronic retroelement in an ion channel gene. Nature *536*, 329-332.

Dobler, S., Petschenka, G., and Pankoke, H. (2011). Coping with toxic plant compounds--the insect's perspective on iridoid glycosides and cardenolides. Phytochemistry *72*, 1593-1604.

Dobzhansky, T. (1937). Genetics and the origin of species (New York: Columbia Univ. Press).

Duffey, R.J., and Goetz, F.W. (1980). The in vitro effects of 17 alpha-hydroxy-20 beta-dihydroprogesterone on germinal vesicle breakdown in brook trout (Salvelinus fontinalis) oocytes. Gen Comp Endocrinol *41*, 563-565.

Earley, E.J., and Jones, C.D. (2011). Next-generation mapping of complex traits with phenotype-based selection and introgression. Genetics *189*, 1203-1209.

Ekker, S.C., Jackson, D.G., von Kessler, D.P., Sun, B.I., Young, K.E., and Beachy, P.A. (1994). The degree of variation in DNA sequence recognition among four *Drosophila* homeotic proteins. Embo J *13*, 3551-3560.

Ewing, A.W., and Bennet-Clark, H.C. (1968). The Courtship Songs of *Drosophila*. Behaviour *31*, 288-301.

Falconer, D.S., and Mackay, T.F.C. (1996). Introduction to quantitative genetics, 4th edn (Essex, England:

Longman).

Furukubo-Tokunaga, K., Flister, S., and Gehring, W.J. (1993). Functional specificity of the Antennapedia homeodomain. Proc Natl Acad Sci U S A *90*, 6360-6364.

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

Galant, R., and Carroll, S.B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. Nature *415*, 910-913.

Garrigan, D., Kingan, S.B., Geneva, A.J., Andolfatto, P., Clark, A.G., Thornton, K.R., and Presgraves, D.C. (2012). Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res *22*, 1499-1511.

Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994). Homeodomain-DNA recognition. Cell *78*, 211-223.

Gellon, G., and McGinnis, W. (1998). Shaping animal body plans in development and evolution by modulation of Hox expression patterns. Bioessays *20*, 116-125.

Georges, M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. Annu Rev Genomics Hum Genet *8*, 131-162.

Goldschmidt, R. (1940). The material basis of evolution (New Have London: Yale University Press ; H. Milford : Oxford University Press).

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

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

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

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

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E.E., Lints, T., Jarvis, E.D., and Scharff, C. (2004). *FoxP2* expression in avian vocal learners and non-learners. J Neurosci *24*, 3164-3175.

Hinaux, H., Bachem, K., Battistara, M., Rossi, M., Xin, Y., Jaenichen, R., Le Poul, Y., Arnoult, L., Kobler, J.M., Grunwald Kadow, I.C.*, et al.* (2018). Revisiting the developmental and cellular role of the pigmentation gene *yellow* in *Drosophila* using a tagged allele. Dev Biol *438*, 111-123.

Hirth, F., Hartmann, B., and Reichert, H. (1998). Homeotic gene action in embryonic brain development of *Drosophila*. Development *125*, 1579-1589.

Hopmann, R., Duncan, D., and Duncan, I. (1995). Transvection in the *iab-5*,*6*,*7* region of the bithorax complex of *Drosophila*: homology independent interactions in *trans*. Genetics *139*, 815-833.

Horn, C., Jaunich, B., and Wimmer, E.A. (2000). Highly sensitive, fluorescent transformation marker for *Drosophila* transgenesis. Dev Genes Evol *210*, 623-629.

Hosemann, K.E., Colosimo, P.F., Brian, R.S., and Kingsley, D.M. (2004). A Simple and Efficient Microinjection Protocol for Making Transgenic Sticklebacks. Behaviour *141*, 1345-1355.

Hu, N., and Castelli-Gair, J. (1999). Study of the posterior spiracles of *Drosophila* as a model to understand the genetic and cellular mechanisms controlling morphogenesis. Dev Biol *214*, 197-210.

Hughes, C.L., and Kaufman, T.C. (2002). Exploring the myriapod body plan: expression patterns of the ten Hox genes in a centipede. Development *129*, 1225-1238.

Indjeian, V.B., Kingman, G.A., Jones, F.C., Guenther, C.A., Grimwood, J., Schmutz, J., Myers, R.M., and Kingsley, D.M. (2016). Evolving new skeletal traits by *cis*-regulatory changes in bone morphogenetic proteins. Cell *164*, 45-56.

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

Jeong, S., Rokas, A., and Carroll, S.B. (2006). Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in *Drosophila* evolution. Cell *125*, 1387-1399.

Johnson, W.C., Ordway, A.J., Watada, M., Pruitt, J.N., Williams, T.M., and Rebeiz, M. (2015). Genetic changes to a transcriptional silencer element confers phenotypic diversity within and between *Drosophila* species. Plos Genet *11*.

Kalay, G., and Wittkopp, P.J. (2010). Nomadic enhancers: tissue-specific *cis*-regulatory elements of *yellow* have divergent genomic positions among *Drosophila* species. Plos Genet *6*, e1001222.

Kapitonov, V.V., and Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. Trends Genet *23*, 521-529.

Karch, F., Galloni, M., Sipos, L., Gausz, J., Gyurkovics, H., and Schedl, P. (1994). *Mcp* and *Fab-7*: molecular analysis of putative boundaries of *cis*-regulatory domains in the bithorax complex of *Drosophila melanogaster*. Nucleic Acids Res *22*, 3138-3146.

Kenyon, C., and Wang, B. (1991). A cluster of Antennapedia-class homeobox genes in a nonsegmented animal. Science *253*, 516-517.

Khila, A., Abouheif, E., and Rowe, L. (2009). Evolution of a novel appendage ground plan in water striders is driven by changes in the *Hox* gene *Ultrabithorax*. Plos Genet *5*, e1000583.

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

Kopp, A., and Duncan, I. (1997). Control of cell fate and polarity in the adult abdominal segments of *Drosophila* by *optomotor-blind*. Development *124*, 3715-3726.

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

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

Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissieres, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A.*, et al.* (2016). Progressive loss of function in a limb enhancer during snake evolution. Cell *167*, 633-642 e611.

Lachaise, D., Harry, M., Solignac, M., Lemeunier, F., Benassi, V., and Cariou, M.L. (2000). Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São Tomé. Proc Biol Sci *267*, 1487-1495.

Lai, C.S., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. Nature *413*, 519-523.

Lewis, E.B. (1964). in The role of chromosomes in development (ed. Locke, M.) (New York: Academic Press).

Lewis, E.B. (1978). A gene complex controlling segmentation in *Drosophila*. Nature *276*, 565-570.

Li-Kroeger, D., Witt, L.M., Grimes, H.L., Cook, T.A., and Gebelein, B. (2008). Hox and senseless antagonism functions as a molecular switch to regulate EGF secretion in the *Drosophila* PNS. Dev Cell *15*, 298-308.

Lin, L., and McGinnis, W. (1992). Mapping functional specificity in the Dfd and Ubx homeo domains. Genes Dev *6*, 1071-1081.

Liu, J., Lemonds, T.R., Marden, J.H., and Popadic, A. (2016). A pathway analysis of melanin patterning in a hemimetabolous insect. Genetics *203*, 403-413.

Mackay, T.F., Stone, E.A., and Ayroles, J.F. (2009). The genetics of quantitative traits: challenges and prospects. Nat Rev Genet *10*, 565-577.

Maeda, R.K., and Karch, F. (2006). The ABC of the BX-C: the bithorax complex explained. Development *133*, 1413-1422.

Mahfooz, N.S., Li, H., and Popadic, A. (2004). Differential expression patterns of the hox gene are associated with differential growth of insect hind legs. Proc Natl Acad Sci U S A *101*, 4877-4882.

Malicki, J., Schughart, K., and McGinnis, W. (1990). Mouse *Hox-2.2* specifies thoracic segmental identity in Drosophila embryos and larvae. Cell *63*, 961-967.

Manak, J.R., and Scott, M.P. (1994). A class act: conservation of homeodomain protein functions. Dev Suppl, 61-77.

Martin, A., Serano, J.M., Jarvis, E., Bruce, H.S., Wang, J., Ray, S., Barker, C.A., O'Connell, L.C., and Patel, N.H. (2016). CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in Crustacean limb specification and evolution. Curr Biol *26*, 14-26.

Matute, D.R., Butler, I.A., and Coyne, J.A. (2009). Little effect of the *tan* locus on pigmentation in female hybrids between *Drosophila santomea* and *D. melanogaster*. Cell *139*, 1180-1188.

McGinnis, N., Kuziora, M.A., and McGinnis, W. (1990). Human *Hox-4.2* and Drosophila *deformed* encode similar regulatory specificities in Drosophila embryos and larvae. Cell *63*, 969-976.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A., and Gehring, W.J. (1984). A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell *37*, 403-408.

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. Cell *68*, 283-302.

Merabet, S., Sambrani, N., Pradel, J., and Graba, Y. (2010). Hox genes : studies from the 20th to the 21st century (ed: JS Deutsch) (New York Austin, Tex.: Springer Science+Business Media; Landes Bioscience).

Mihaly, J., Barges, S., Sipos, L., Maeda, R., Cleard, F., Hogga, I., Bender, W., Gyurkovics, H., and Karch,

F. (2006). Dissecting the regulatory landscape of the *Abd-B* gene of the bithorax complex. Development *133*, 2983-2993.

Morata, G., and Garcia-Bellido, A. (1976). Developmental analysis of some mutants of the bithorax system of *Drosophila*. Wilehm Roux Arch Dev Biol *179*, 125-143.

Mundy, N.I. (2005). A window on the genetics of evolution: MC1R and plumage colouration in birds. Proc Biol Sci *272*, 1633-1640.

Nishida, R. (2002). Sequestration of defensive substances from plants by Lepidoptera. Annu Rev Entomol *47*, 57-92.

Obbard, D.J., Maclennan, J., Kim, K.W., Rambaut, A., O'Grady, P.M., and Jiggins, F.M. (2012). Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. Mol Biol Evol *29*, 3459-3473.

Ordway, A.J., Hancuch, K.N., Johnson, W., Wiliams, T.M., and Rebeiz, M. (2014). The expansion of body coloration involves coordinated evolution in cis and trans within the pigmentation regulatory network of *Drosophila prostipennis*. Dev Biol *392*, 431-440.

Parker, H.J., Pushel, I., and Krumlauf, R. (2018). Coupling the roles of Hox genes to regulatory networks patterning cranial neural crest. Dev Biol.

Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A., and Bejerano, G. (2013). Enhancers: five essential questions. Nat Rev Genet *14*, 288-295.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970-985.

Peter, I.S., and Davidson, E.H. (2015). Genomic control process: development and evolution (London, UK ; San Diego, CA, USA: Academic Press is an imprint of Elsevier).

Pool, J.E., and Aquadro, C.F. (2007). The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. Mol Ecol *16*, 2844-2851.

Prescott, S.L., Srinivasan, R., Marchetto, M.C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F.H., Swigut, T., and Wysocka, J. (2015). Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. Cell *163*, 68-83.

Protas, M.E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W.R., Zon, L.I., Borowsky, R., and Tabin, C.J. (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet *38*, 107-111.

Raff, R.A. (1996). The shape of life: genes, development, and the evolution of animal form (Chicago: University of Chicago Press).

Raff, R.A. (2000). Evo-devo: the evolution of a new discipline. Nat Rev Genet *1*, 74-79.

Rebeiz, M., Patel, N.H., and Hinman, V.F. (2015). Unraveling the tangled skein: The evolution of transcriptional regulatory networks in development. Annu Rev Genomics Hum Genet *16*, 103-131.

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

Rebeiz, M., and Posakony, J.W. (2004). GenePalette: a universal software tool for genome sequence visualization and analysis. Dev Biol *271*, 431-438.

Rebeiz, M., Ramos-Womack, M., Jeong, S., Andolfatto, P., Werner, T., True, J., Stern, D.L., and Carroll,

S.B. (2009b). Evolution of the *tan* locus contributed to pigment loss in *Drosophila santomea*: A response to Matute et al. Cell *139*, 1189-1196.

Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R.*, et al.* (2011). *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. Science *333*, 1137.

Reimchen, T. (1983). Structural relationships between spines and lateral plates in threespine stickleback (*Gasterosteus aculeatus*), Vol 37.

Reznick, D.N., and Ricklefs, R.E. (2009). Darwin's bridge between microevolution and macroevolution. Nature *457*, 837-842.

Roeske, M.J., Camino, E.M., Grover, S., Rebeiz, M., and Williams, T.M. (2018). Cis-regulatory evolution integrated the Bric-à-brac transcription factors into a novel fruit fly gene regulatory network. Elife *7*.

Rogers, B.T., Peterson, M.D., and Kaufman, T.C. (1997). Evolution of the insect body plan as revealed by the *Sex combs reduced* expression pattern. Development *124*, 149-157.

Rogers, W.A., Grover, S., Stringer, S.J., Parks, J., Rebeiz, M., and Williams, T.M. (2014). A survey of the *trans*-regulatory landscape for *Drosophila melanogaster* abdominal pigmentation. Dev Biol *385*, 417-432.

Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. Nature *415*, 914-917.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., and Kingsley, D.M. (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature *428*, 717-723.

Smith, A.F., Posakony, J.W., and Rebeiz, M. (2017). Automated tools for comparative sequence analysis of genic regions using the GenePalette application. Dev Biol *429*, 158-164.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70-71.

Stearns, F.W. (2010). One hundred years of pleiotropy: a retrospective. Genetics *186*, 767-773.

Stern, D.L. (1998). A role of *Ultrabithorax* in morphological differences between *Drosophila* species. Nature *396*, 463-466.

Stern, D.L. (2000). Evolutionary developmental biology and the problem of variation. Evolution *54*, 1079-1091.

Stern, D.L. (2014). Identification of loci that cause phenotypic variation in diverse species with the reciprocal hemizygosity test. Trends Genet *30*, 547-554.

Struhl, G., Barbash, D.A., and Lawrence, P.A. (1997). Hedgehog organises the pattern and polarity of epidermal cells in the *Drosophila* abdomen. Development *124*, 2143-2154.

Swalla, B.J., and Jeffery, W.R. (1996). Requirement of the *Manx* gene for expression of chordate features in a tailless ascidian larva. Science *274*, 1205-1208.

Tichy, A.L., Ray, A., and Carlson, J.R. (2008). A new *Drosophila* POU gene, *pdm3*, acts in odor receptor

expression and axon targeting of olfactory neurons. J Neurosci *28*, 7121-7129.

True, J.R., Yeh, S.D., Hovemann, B.T., Kemme, T., Meinertzhagen, I.A., Edwards, T.N., Liou, S.R., Han, Q., and Li, J. (2005). *Drosophila tan* encodes a novel hydrolase required in pigmentation and vision. Plos Genet *1*, e63.

Valentine, J.W. (2004). On the origin of phyla (Chicago: University of Chicago Press).

Van Auken, K., Weaver, D.C., Edgar, L.G., and Wood, W.B. (2000). *Caenorhabditis elegans* embryonic axial patterning requires two recently discovered posterior-group Hox genes. Proc Natl Acad Sci U S A *97*, 4499-4503.

Vlad, D., Kierzkowski, D., Rast, M.I., Vuolo, F., Dello Ioio, R., Galinha, C., Gan, X., Hajheidari, M., Hay, A., Smith, R.S.*, et al.* (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. Science *343*, 780-783.

Vuolo, F., Mentink, R.A., Hajheidari, M., Bailey, C.D., Filatov, D.A., and Tsiantis, M. (2016). Coupled enhancer and coding sequence evolution of a homeobox gene shaped leaf diversity. Genes Dev *30*, 2370-2375.

Walter, M.F., Black, B.C., Afshar, G., Kermabon, A.Y., Wright, T.R., and Biessmann, H. (1991). Temporal and spatial expression of the *yellow* gene in correlation with cuticle formation and dopa decarboxylase activity in *Drosophila* development. Dev Biol *147*, 32-45.

Warren, R.W., Nagy, L., Selegue, J., Gates, J., and Carroll, S. (1994). Evolution of homeotic gene regulation and function in flies and butterflies. Nature *372*, 458-461.

Washburn, S.L. (1963). Classification and human evolution (Chicago,: Aldine Pub. Co.).

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Weatherbee, S. D., Halder, G., Kim, J., Hudson, A., and Carroll, S. (1998). Ultrabithorax regulates genes at several levels of the wing-patterning hierarchy to shape the development of the *Drosophila* haltere. Genes Dev, *12*, 1474–1482.

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

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

Wittkopp, P.J., and Kalay, G. (2011). *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet *13*, 59-69.

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

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

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

Wright, T.R. (1996). The Wilhelmine E. Key 1992 Invitational lecture. Phenotypic analysis of the Dopa decarboxylase gene cluster mutants in *Drosophila melanogaster*. J Hered *87*, 175-190.

Yassin, A., Delaney, E.K., Reddiex, A.J., Seher, T.D., Bastide, H., Appleton, N.C., Lack, J.B., David, J.R., Chenoweth, S.F., Pool, J.E.*, et al.* (2016). The *pdm3* locus is a hotspot for recurrent evolution of female-limited color dimorphism in *Drosophila*. Curr Biol *26*, 2412-2422.

Yoder, J.H., and Carroll, S.B. (2006). The evolution of abdominal reduction and the recent origin of distinct Abdominal-B transcript classes in Diptera. Evol Dev *8*, 241-251.

Yu, Z.S., Ren, M.D., Wang, Z.X., Zhang, B., Rong, Y.K.S., Jiao, R.J., and Gao, G.J. (2013). Highly efficient genome modifications mediated by CRISPR/Cas9 in *Drosophila*. Genetics *195*, 289-+.

Zeng, W., Andrew, D.J., Mathies, L.D., Horner, M.A., and Scott, M.P. (1993). Ectopic expression and function of the *Antp* and *Scr* homeotic genes: the N terminus of the homeodomain is critical to functional specificity. Development *118*, 339-352.

Zhen, Y., Aardema, M.L., Medina, E.M., Schumer, M., and Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community. Science *337*, 1634-1637.